March 2018

Analog Computing using 1T1R Crossbar Arrays

Yunning Li
University of Massachusetts Amherst

ANALOG COMPUTING USING 1T1R CROSSBAR ARRAYS

A Thesis Presented

by

YUNNING LI

A thesis submitted to the University of Massachusetts Amherst in partial fulfilment of the requirement of the degree of

MASTER OF SCIENCE IN ELECTRICAL AND COMPUTER ENGENEERING

FEBRUARY 2018

Electrical and Computer Engineering

ANALOG COMPUTING USING 1T1R CROSSBAR ARRAYS

A Thesis Presented

by

YUNNING LI

Approved as to style and content by:

_____

Professor Joshua J. Yang, Chair

_____

Associate Professor Qiangfei Xia

_____

Assistant Professor Dan Holcomb

_____

C. V. Hollot, Department Head

Electrical & Computer Engineering

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to all those who made this thesis possible. First and foremost, I would like to express my sincere gratitude to my advisor, Professor Joshua (Jianhua) Yang. His passion and enthusiasm in research has greatly impressed me. It is my great pleasure to be a student of him. His valuable guidance and constant support are crucial to the advancements of these projects. He has been a source of inspiration and knowledge throughout my years at UMass. He always encourages me to explore new ideas, and I truly appreciate the degrees of freedom that he offered on conducting researches. Without his unselfish support, I cannot achieve my success so far.

I wish to thank Professor Qiangfei Xia, for the collaboration during my recent years at UMass. He is very kind and always willing to help but sharp at key issues in the research. His great ideas, attitude and foresight have deeply influenced and inspired me.

I also would like to thank Professor Dan Holcomb for serving as my dissertation committee member, reading the dissertation and providing valuable comments to this thesis.

I owe many thanks to all current and former members from our research team or professor Xia's group: Dr. Zhongrui Wang, Dr. Saumil Joshi, Dr. Jung Ho Yoon, Dr. Yibo Li, Mr. Peng Yan, Mr. Wenhao Song, Mr. Mingyi Rao, Mr. Rivu Midya, Mr. Navnidhi Kumar Upadhyay, Mr. Ye Zhuo, Mr. Shiva Asapu, Dr. Peng Lin, Dr. Hao Jiang, Dr. Shuang Pi, Mr. Can Li, Ms. Siyan Lin, and Ms. Rui Zhang, for their generous help, valuable discussions and kind encouragement. They are irreplaceable sources of thoughts and support, and I have really learnt a lot from them in these years. Without their help and support, I could never finish my master's degree.

My deepest gratitude is with my family members: my grandfather, my father and mother, my uncles and aunts and my brothers and sisters. They always encourage me to pursue my dream and their support is the strongest driving force for me to go forward.

Last but not least, I would like to thank my roommates and all friends for their support within these years. Your encouragements have kept me up in the hardest time.

ABSTRACT

ANALOG COMPUTING USING 1T1R CROSSBAR ARRAYS

FEBRUARY 2018

YUNNING LI

M.S.E.C.E., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Jianhua Joshua Yang

Memristor is a novel passive electronic device which has attracted increasing attention in the past decade as a promising candidate for new generation non-volatile memory and analog computing. Memristor is a two-terminal component having the structure of metal-insulator-metal (MIM). It possesses a unique electrical behavior, which is, that the resistance state can be reconfigured electrically and maintained after the removal of the applied external bias. The resistance hence depends on the history of the applied voltage and conducted current. Compared with conventional transistors, memristor exhibits simple structure, better scalability, and rich electrical behaviors, which may enable novel electronic applications in data storage, analog computing, neuromorphic applications, circuits design etc.

This work focuses on three main topics regarding using memristor crossbars for analog computing. The first is about building customized measurement systems for different circuit applications. There are two measurement systems will be introduced. One is a switches-matrix that enables the measurement of large arrays of devices. Another measurement system is dedicated for 1T1R crossbar arrays. The system features multi-channel concurrent reading and driving, which could be used for analog and parallel computing measurements. The second topic is using memristors as a dot product engine

(DPE) for the computation accelerating in image processing. The intrinsic advantage of memristor crossbars is that it allows for a fast and energy efficient way of performing VMM operations with analog inputs directly. This is because the conductance of memristors encodes the weight of linear transformation in the computing, and parallel multiplications take place at all the crossing points of the entire crossbar simultaneously. The third one is to use memristors in reconfigurable analog circuits design, also termed as memFPAA. A common belief is that if data could be processed by analog circuits directly without digitalizing them first, then the time and power consumption could be substantially reduced. Designing specific analog computing circuits has been expensive and time-consuming because of the inconvenience in prototyping the analog circuit design. FPAA provides a potential approach to explore the possibility of developing analog circuits on a reconfigurable platform.

Keywords: measurement system, DPE, image compression, convolution, FPAA, audio equalizer, pattern classification, memristor, 1T1R.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION TO MEMRISTOR

1.1 Background

The new technological revolution since 20th century has changed human life significantly. Electronic devices have been widely used in our daily life, such as computer, smartphones. In addition, we are entering the IoT (internet of things) era, which has put enormous demand on data storage and transmission. Current commercial memory technology still focusses on conventional complementary metal-oxide-semiconductor (CMOS), such as the widely used FLASH or SRAM and DRARM. As CMOS scaling is approaching its physical limit and Moore's law will come to an end in a decade, researching into novel electronic devices is of great importance.

Varieties of emerging devices for memory have been explored, including phase change random access memory (PCRAM), spin-transfer-torque magnetic random-access memory (STTMRAM), and ferroelectric RAM (FeRAM). PCRAM uses materials' phase change between amorphous and crystalline induced by Joule heating to represent device's state, amorphous state with high resistance and crystalline state with low resistance. Tuning applied voltage amplitude and duration can switch device's state from amorphous to crystalline and vice versa. STTMRAM is with a structure of an insulator layer sandwiched between two magnetic layers. Typically, if the two magnetic layers have same spin polarization, device is at low resistance state (LRS), conversely, showing high resistance state (HRS). The spin polarization of magnetic layer can be transferred via applying sufficient current of different polarities. FeRAM has a similar structure as a

DRAM but using a ferroelectric layer. The low power consumption, fast speed and superior endurance makes it a potential device that is widely used in memory.

Another promising candidate as a next generation non-volatile memory is resistive random-access memory (ReRAM), which is based on resistance switching behavior.

1.2 Memristive Devices

Here we demonstrate a novel nano-device with extremely simple structure, reconfigurable two-terminal electronic circuit element – memristor. Memristor is one of the most promising nano-devices which can potentially replace CMOS devices for advanced computing including digital and analog circuit applications. Few decades ago, the idea of memristor (or memristive) device was first-time introduced by Leon Chua. He proposed the existence of the fourth passive fundamental circuit element, with resistor, capacitor, and inductor. At the beginning, the device was proposed as a missing element that links properties between magnetic flux and electric charge, as the equation in Fig. 1.1 shows, accumulation of flux of magnetic or charge leads to the resistance changing. The first claim of the invention of a memristor was by Hewlett-Packard Lab in 2008, which provided the first practical demonstration, researchers observed a memristive behavior in a thin-film titanium dioxide insulator layer.

An essential model of memristor can be understood as a two-terminal device based on a metal/insulator/metal (MIM) structure. In general, the structure is illustrated in Fig. 1.2, a device consists of a hafnium oxide or other material layer sandwiched by two platinum electrodes. In real circuit, device is made in an elementary crossbar or cross-

point form, also through sharing electrodes it can be built into a memristor array. Albeit the simple structure, memristors exhibits fascinating electrical properties such as high speed and low power consumption of switching, great scalability, and cycling ability.

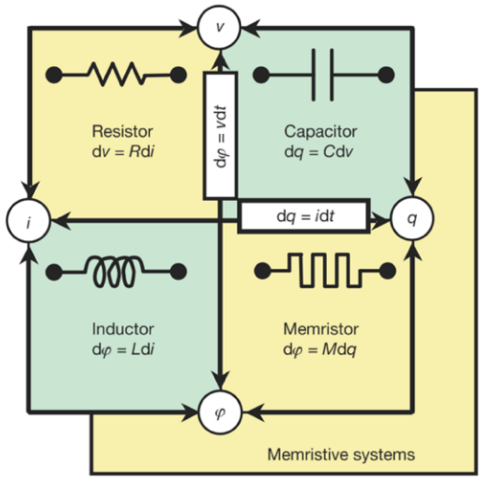


Figure 1.1 : The four fundamental two-terminal passive electric elements: Resistor, Capacitor, Inductor, Memristor. As the equation illustrated, memristor combines the magnetic and electric charge. For example, the memristor state is the relation of flux of charge through it.

The main characteristic of memristor is the resistance of device can be adjusted by applying an external voltage. The common explanation for resistive switching of memristor is the formation and rapturing of conductive filaments (channels) in the metal-oxide thin layer, which lead the state of device shift from the “OFF” state to the “ON” state or the opposite process. The existence of filaments in the switching layer between top electrode and bottom electrode creates a low-resistance state (LRS), while the breaking of filaments recovers the device to high-resistance state (HRS). Figure 1.3 gives the specific processes of filament forming and breakdown, (a) the virgin state without filaments exist in the switching layer, device presents HRS, it could over few Giga Ohm; (b) during the electroforming process, conductive elements are generated and drift in the insulator; (c) conductive filament is formed, device presents the on state (HRS to LRS

transition). (d) backward voltage inverses the process of filament form leads device from LRS back to HRS.

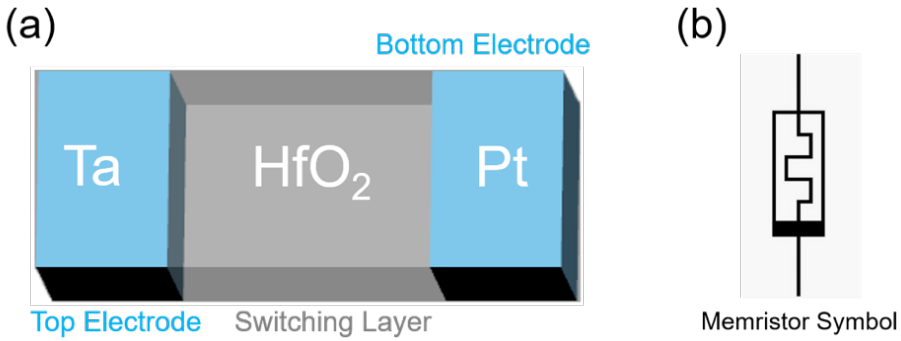


Figure 1.2 : Structure of memristor device with Metal/Insulator/Metal configurations.

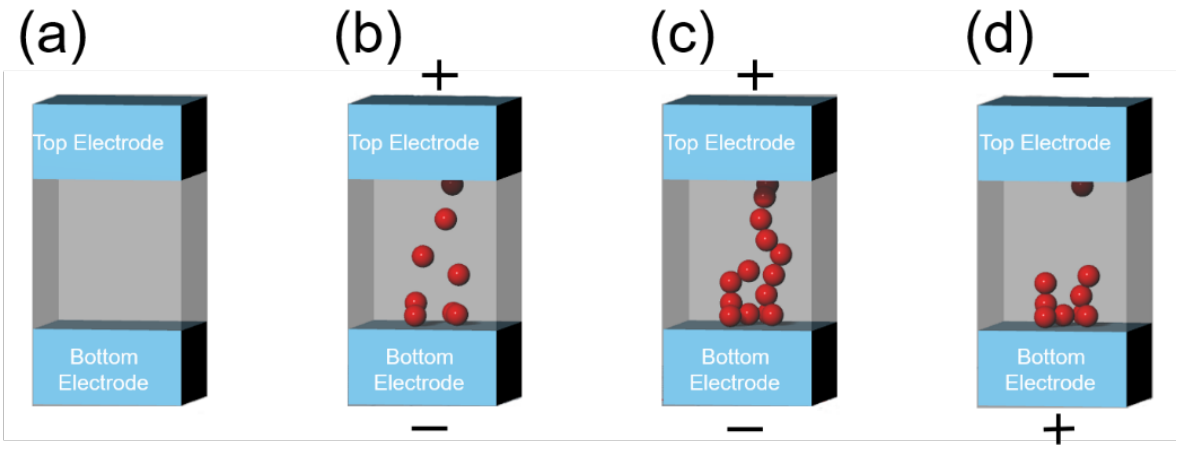


Figure 1.3 : Schematic of resistive switching according to a conductive filaments mechanism. a. Virgin state presents HRS. b. Electroforming process, conductive channels are being created. c. Conductive filaments are formed, (ON state, LRS). d. Filament breakdown in an inverse voltage (OFF state, HRS).

The resistive switching or filaments formation mechanisms have been widely discussed, there are two suggested explanations involve the valence change memory (VCM) and the electrochemical metallization memory (ECM). The VCM builds on induced anion migration that modifies the stoichiometry of the insulator region via oxidation-reduction reactions. The ECM relies on the oxidative interfacial dissolution of an active metal electrode, followed by subsequent cation migration across an ion-

conducting electrolyte layer, acting as an insulator. Generally, the switching behaviors of memristor can be roughly classified into two modes, unipolar switching and bipolar switching. For this work has been done by using bipolar devices, here only discuss the behavior about it. The schematics of I-V curve characteristic of bipolar resistive switching are illustrated in Fig. 1.4, where opposite voltage polarities are necessary. The set operation defines as the movement of resistance from HRS to LRS, which requires a positive voltage. The reset operation is the move from LRS to HRS, negative voltage was applied. Typically, the set process with a compliance current to protect device from burning by Joule heating.

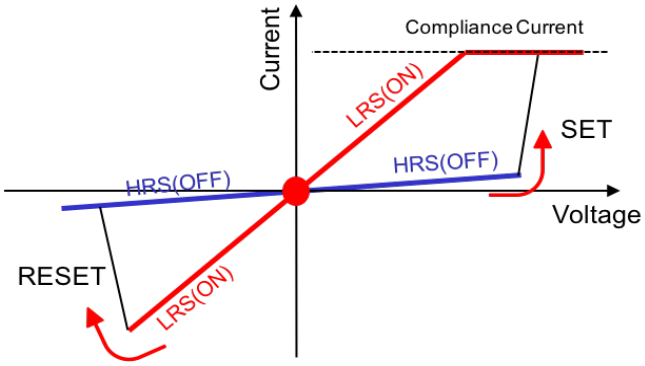


Figure 1.4 : Bipolar switching behavior in memristive devices.

The most promising application of using memristor is as a non-volatile memory. Table 1 list the key parameters comparison between memristor with some commercial memory technologies. It tells that the memristor has similar density as flash (NAND) and operation speed as dynamic random-access memory (DRAM). The non-volatile property enables zero power consumption for holding a state, which make memristor as an ideal device for use in the next generation computer. A hafnium-oxide based memristor was used in this work, as the Fig. 1.1 shows, this device is fabricated in our clean room, it

uses 20nm thick Pt as the bottom electrode, 50nm Ta as the top electrode and 5nm HfO2 as the switching layer.

Table 1.1 : Comparison between memristor and CMOS based memory technologies.

	Memristor	DRAM	SRAM	Flash(NAND)
Cell Density(F^2)	4	6-30	140	4-6
Cell Element	1R/1T1R	1T1C	6T	1T
Retention	>10years	>64ms	Weeks?	>10years
Endurance	$>10^{12}$ cycles	$>10^{16}$ cycles	$>10^{16}$ cycles	$>10^{5}$ cycles
Read Time	<2ns	2ns	0.2ns	0.1ms
Write Time	<10ns	10~50ns	10~100ns	200us

This device has many advantages including low programming voltage, fast switching speed, high endurance and reliable retention. Its retention and endurance performance are showing in Fig. 1.5a-b, endurance tested under SET pulses of 1.3 V/100 ns and −3.05 V/100 ns for RESET. This device also shows an excellent multi-level performance will be discussed in the 1T1R section.

1.3 1T1R

The simple structure of memristor gives a possibility integrated it on a transistor, this structure also been well known as 1T1R. A typical 1T1R combination as shown in Fig. 1.6a, the bottom electrode of memristor is connected with the extension of the drain terminal of transistor, then memristor can be accessed by using the top electrode and the source terminal, as well as using gate to control the transistor's ON and OFF. Fig. 1.6b is the symbol of a 1T1R device. In this work, transistors are built on 2um feature size by a commercial company, which provides advantages on allowing high operating voltage (could over 10V), and very small on-resistance of transistor.

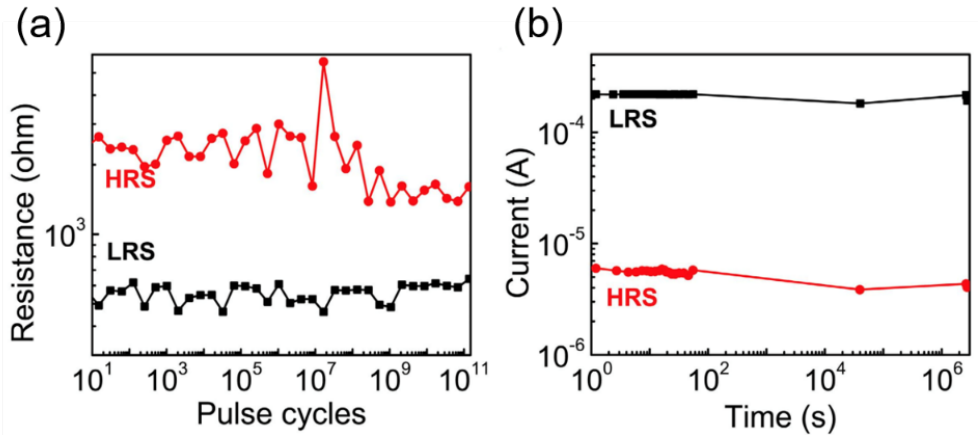


Figure 1.5 : Device endurance and retention test. a. 120 billion switching cycles have been demonstrated with pulses of 1.3 V/100 ns for SET and −3.05 V/100 ns for RESET. b. Retention test at room temperature shows no evident degradation after 1 month.

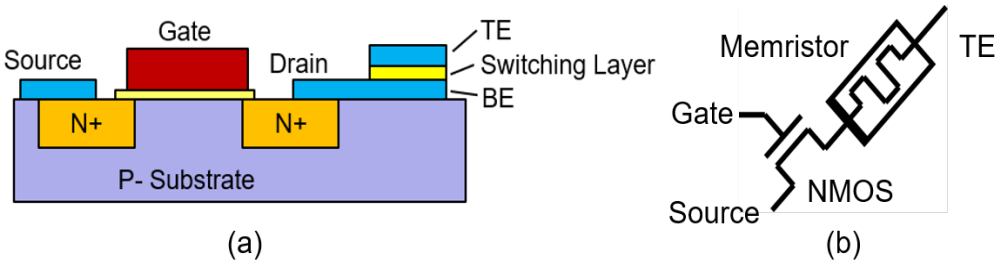


Figure 1.6 : The 1T1R structure and its symbol. a. 1T1R, the combination one N-type transistor and one HfO2 based memristor. b. the common symbol of 1T1R.

Transistor in the combination not only plays the role of access controlling, also can use gate voltage to limit the current through pass the memristor, like the compliance current realized by a semiconductor analyzer. In this case, it can also be used to tune the state of memristor by applying different gate voltage. Fig 1.7 illustrates the performance of transistor, the I-V curve. At 5V gate voltage, the Ids can reach 5mA, that is much higher than the memristor operating current (about 1mA).

The symmetry and linearity of transistor are also important properties that is a guarantee for using memristor in high accuracy computing, such as most analog

computing applications. Fig. 1.8a gives symmetry performance of the transistor, it was measured by applying driving voltage on source terminal and ground the drain; Fig. 1.8b is on-resistance changing at gate voltage of 5V, the slight on-resistance disturbance is negligible in the operational resistance range of memristor.

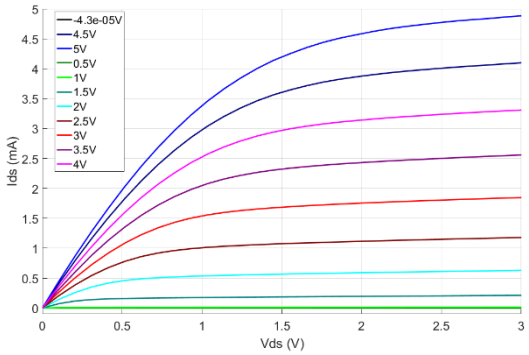


Figure 1.7 : I-V curves of transistor used in 1T1R crossbar array. DC sweep from 0V – 3V, gate voltage from 0V – 5V, 0.5V one step.

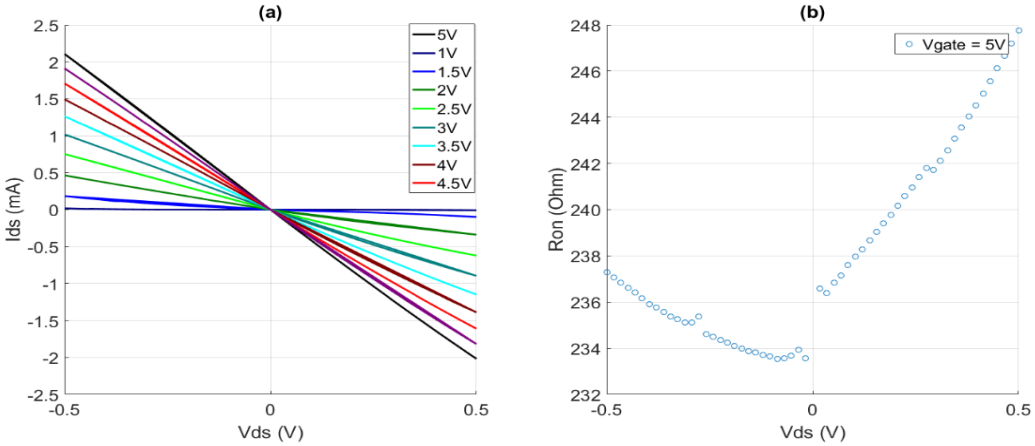


Figure 1.8 : Linearity and on resistance of transistor. a. sweep voltage from -0.5 to 0.5Vapplied on source, gate voltage changing from 0V to 5V, 0.5V a step. b. transistor on-resistance at 5V gate voltage.

Another important benefit of involving transistor is to avoid the sneak path current in a passive array. Sneak path current is an intrinsic challenge for a passive array, as the Fig. 1.9 demonstrated. An applied voltage on target device (M4), current is not

only pass through it (the signal path), but also through other devices, such as the sneak path (M1 – M4). The current level at detected terminal cannot accurately reflect the resistance state of the target device, it is a comprehensive result of the whole array. It could be much worse if the array much bigger. To avoid the sneak path problem, researchers have explored many solutions. For example, using non-linearity device array, in which different device has different voltage drop to limit the sneak path current. Although this solution without complicated passive array structure, but it is in the sacrifices of linear and accurate computing, that limits the application fields of array.

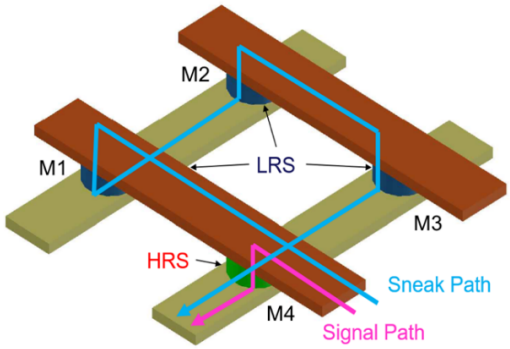


Figure 1.9 : A demonstration of sneak path on a 2x2 passive array. Memristor M1-M3 on LRS and M4 on HRS.

A typical 1T1R crossbar array is illustrated in Fig. 1.10. Top electrodes of a column of devices share a common bit-line and a row of devices' source terminals share a common world-line, as well as a gate-line that is parallel with bit-line connects to all gate terminals in the column. In program mode, turn on transistors on the corresponding column, then apply voltage either on top electrode or source of target device to tune its resistance. For all transistors on other columns on "OFF" state cannot stop the sneak path from influencing the target device. On the other hand, if all transistors are turned on that leads the 1T1R array work like a passive array. Directly for applications, such as vector-matrix multiplication.

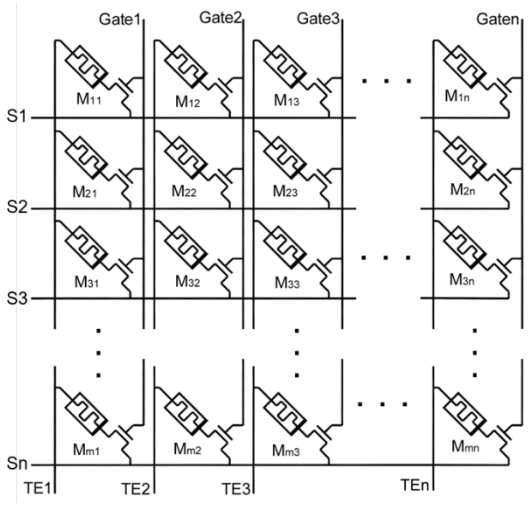
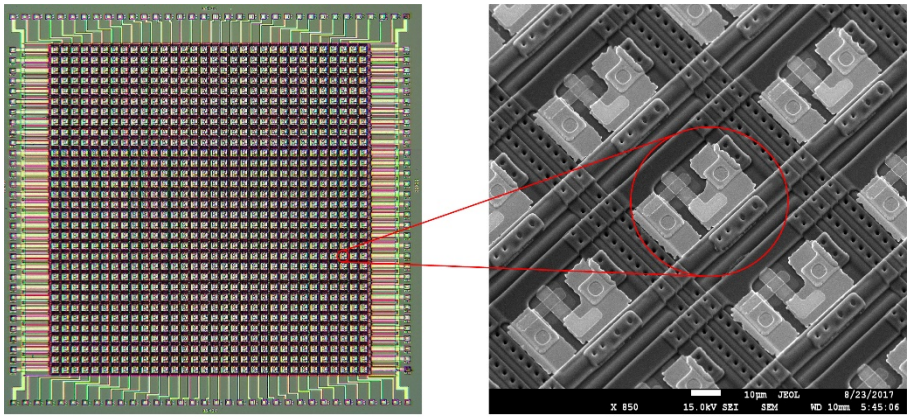


Figure 1.10 : Schematic of 1T1R array. One memristor in series with a transistor which is parallel with other identical structures forms the array.

Fig 1.11a illustrates the real device image. It is a 32x32 1T1R crossbar array under a probe card. In the center region, one dot is a combination of one transistor and a drift memristor. TEM picture is also shown in the figure Fig 1.11b, which gives more details about the device. It can be seen that a memristor crossbar (white part) is built on the top a transistor. Wires are shared with gate, source terminals and top electrodes. Transistors with feature size of 2 μm were used in this work. The transistors were fabricated in a commercial fab, which gives small wire resistance (about 0.7Ω per block). Photolithography, thin film deposition, and liftoff were used to integrate the memristors with the wired transistors. In order to remove native metal oxide layers, argon plasma treatment was applied on the transistor chip which gives a better electrical connection. Metal vias were created by the sputtering of 5-nm silver (Ag) and 200-nm palladium (Pd). This was followed by the lifting off of samples in warm acetone. Samples were annealed at 300 °C in nitrogen ambience (flow 20 sccm) for half an hour. A 60-nm thick Pd bottom electrode was sputtered on a 5-nm tantalum (Ta) adhesive layer. To ensure the high film quality and steps coverage, we used water and tetrakis (dimethylamido)

hafnium as precursors for depositing a 5-nm HfO_2 switching layer by ALD at 250 °C. Photolithography and reactive ion etch (RIE) using CHF_3/O_2 patterned the switching layer. Finally, the top electrode was sputtered and lifted-off with a 50-nm thick Ta and covered by a 10-nm thick Pd as a passivation layer.



(a) (b)

Figure 1.11 : Photographed details of 1T1R crossbar array and TEM. a. A picture of 32x32 1T1R array taken by a microscopy camera. b. SEM photo shows the details of 1T1R array, the cross section in the center is a memristor built on the top of a transistor.

Multiple resistance states have been achived for the device by using diferent gate voltages. As shown in Fig. 1.12, the device was set to 28 resistance states with DC sweeps (0V ~ 2V ~ 0V) by controlling the gate voltage from 0V to 2.5V. A higher gate voltage led to a higher current pass through the device which set the device to lower resistance states. To demonstrate the gate controlling can achieve programming of resistance states, negative sweep voltage was applied to reset device to high resistance state after each positive loop.

As the Fig. 1.8 illustrated the linearity of transistor, Fig. 1.13 is showing the linearity of 1T1R device, in which the memristor was set to different resistance levels cover the range of 1k to 10k. the excellent linear performance is one of the guarantees of linear analog computing.

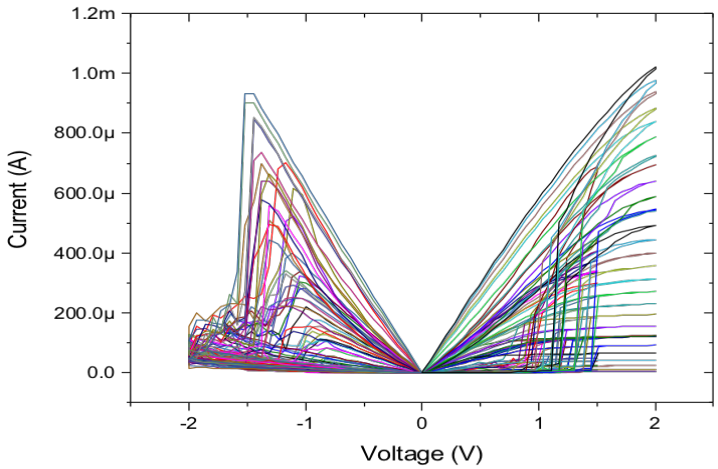


Figure 1.12 : Demonstration of using transistor tuning conductance state in 1T1R. The device can be programmed to lower resistance level by increasing gate voltage a little, the gate voltage is swept from 0.8V to 2.5V.

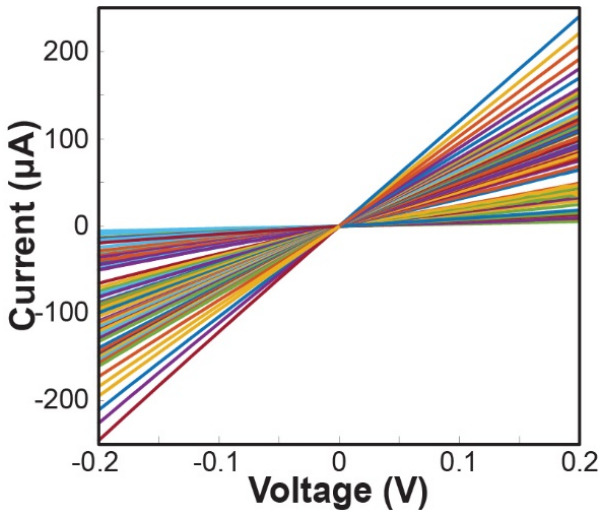


Figure 1.13 : Linearity of 1T1R device. DC voltage sweeps of different memristor conductance states (100µS ~ 1000µS), showing good I-V linearity of memristors.

1.4 Summary

The requirements of high density, high speed, low power consumption stimulate us to pursue new technologies, especially we are facing the bottleneck of CMOS technology now and predictably in the near future. There are several novel devices that have been explored, including PCRAM, STTMRAM, FeRAM. All of them illustrate unique characteristics and wide applications. Memristive devices are promising

candidates for next generation computing architecture and data storage. It is a simple two-terminal passive device shows many distinct properties, such as scalability, multi-conductance levels, high write and read speed, and low power consumption, and also CMOS compatibility. Memristors cannot only be used for data storage, but also for computing, such as analog computing or neuromorphic computing. Although, there are still crucial challenges in the development of memristive devices, with the increasing understating of its structure and mechanism, the performance of memristors could be further improved.

1.5 Reference

1 L. Chua, *et al.* Memristor-The missing circuit element. *IEEE Transactions on Circuit Theory* 18, 507-519 (1971).

2 P. M. Sheridan, *et al.* Sparse coding with memristor networks. *Nat Nanotechnol* **12**, 784-789, (2017).

3 D. B. Strukov, *et al.* The missing memristor found. *Nature* 453, 80-83 (2008).

4 J. J. Yang, *et al.* Memristive switching mechanism for metal/oxide/metal nano-devices. *Nat Nanotechnol* 3, 429 (2008).

5 Q. Xia, *et al.* Memristor−CMOS Hybrid Integrated Circuits for Reconfigurable Logic. *Nano Letters* 9, 3640-3645 (2009).

6 J. J. Yang, *et al.* Memristive devices for computing. *Nat Nanotechnol* 8, 13-24 (2013).

7 H. Jiang, *et al.* Sub-10 nm Ta Channel Responsible for Superior Performance of a HfO2 Memristor. *Sci Rep* 6, 28525 (2016).

8 L. Gao, *et al.* Demonstration of Convolution Kernel Operation on Resistive Cross-Point Array. *IEEE Electron Device Letters* 37, 870-873 (2016).

9 M. Hu, *et al.* in *2016 IEEE International Conference on Rebooting Computing (ICRC).* 1-5 (San Diego, CA, USA, 2016).

10 H. Wu, *et al.* Resistive Random Access Memory for Future Information Processing System. in *Proceedings of the IEEE* PP-99 (2017).

11 B. Chakrabarti, *et al.* A multiply-add engine with monolithically integrated 3D memristor crossbar/CMOS hybrid circuit. *Sci Rep* 7, 42429 (2017).

12 C. Li, *et al.* Analog signal and image processing with large memristor crossbars. *Nature electronics,* (2017 accepted).

13 P. G. Gulak, *et al.* in *Microelectronics and VLSI, 1995. TENCON '95., IEEE Region 10 International Conference on.* 123-126.

14 M. Hu, *et al.* The Dot Product Engine: a memristor-based experimental platform for computational acceleration. (2017, in preparation).

15 H. Kutuk, *et al.* A field-programmable analog array (FPAA) using switched-capacitor techniques in circuits and Systems ISCAS (1996).

CHAPTER 2

MEASUREMENT SETUP

2.1 Conventional Measurement Setup

The characterization of memristive devices requires a unique setup to do the electrical tests, which can reveal their properties over different time scale, from fast pulse response to data retention and endurance. For this purpose, a setup based on the B1500A semiconductor analyzer from Keysight Technologies could implement DC and pulse measurement is preferred. For conventional DC and Pulse measurement, the B1500A semiconductor analyzer installs the following plug-in modules:

1. high-resolution source measurement units (HRSMU) with a maximum current and voltage of ±100 mA and ±100 V, respectively, and a current and voltage mean resolution of 0.1 fA and 0.5 μV, respectively, and a current and voltage mean resolution of 0.1 fA and 0.5 μV, respectively.
2. waveform generation/fast measurement unit (WGFMU) with a minimum timing resolution of 10 ns and a minimum measurement repetition rate of 5 ns. The maximum voltage range is 10 V (± 5 V or 0 – 10 V) with a variable current measurement ranges from 1 μA to 10 mA.

Those modules all can be switched between the measurement of current or voltage. Almost all measurement parameters could be customized, for instance the voltage amplitude, steps, rise/fall time and delay etc. The general setup for a memristor characterization measurement is as shown in Figure 2.1. The device-under-test (DUT) is on top of the chuck of a probe station. Probe tips with tri-axial cables connected to the

B1500A are landed on the devices. Depends on the type of measurement, either DC or pulse module is employed.

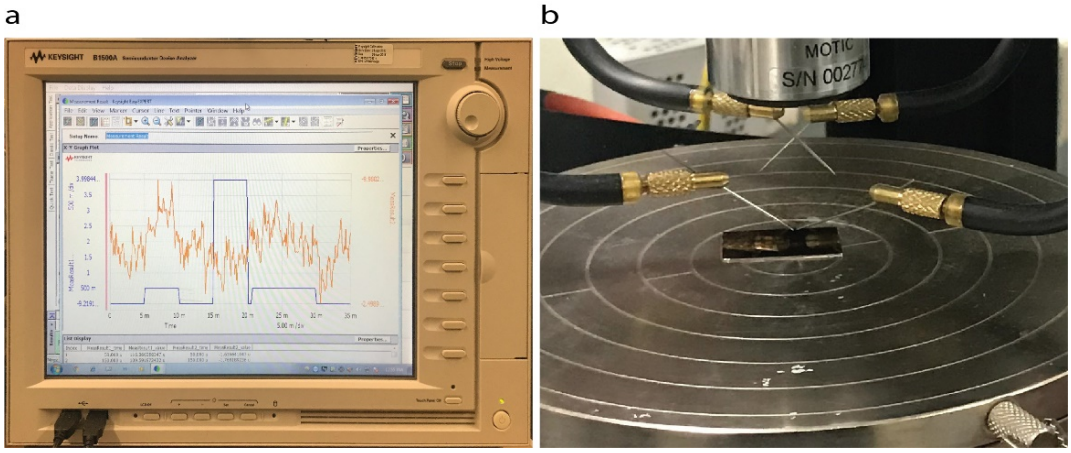


Figure 2.1 : Measurement setup utilizing B1500A and a probe station with few probe tips enables DC and pulses measurements. A fast pulse was applied on device, pulses after and before the high amplitude pulse are reading pulses.

A custom C# based program was developed to accompany with the setup for necessary functionality measurements, such as DC voltage sweep, compliance current, pulse endurance and retention. The program can do operations for memristor includes FORM, RESET, SET and READ on both DC and pulse modes. Also, it can be used to collect measurement data for images plots or further data analysis.

2.2 Switching-Matrix

As our understanding on memristor switching mechanism and electronic properties is deepening, more requirements are being placed on measurement instruments for exploring more possibilities of using the device. Also, researchers are not satisfied anymore on studying a single device. However, the commercial measurement instruments do not come up with the researching requirements and unique application demands, such as large scale of device statistics and parallel computing. In order to overcome shortages

of test tools, we built a system which can help the measurement on the large-scale and stochastic analysis of devices. As the Figure 2.2 shows this system is a switching-matrix, that works in conjunction with the B1500A semiconductor analyzer. In the rough diagram, 1T1R crossbar array is the DUT, each access terminal connects with a switch, and the common terminals of switches are connected to the B1500A. To turn on, corresponding switches can access any single device in the array, for example, applying voltage on Gate1, BE1 and TE1 that is accessing the device M_{11}. The system also includes a controlling board which can be used to control switches by commands from a computer, details of this board are not shown the figure.

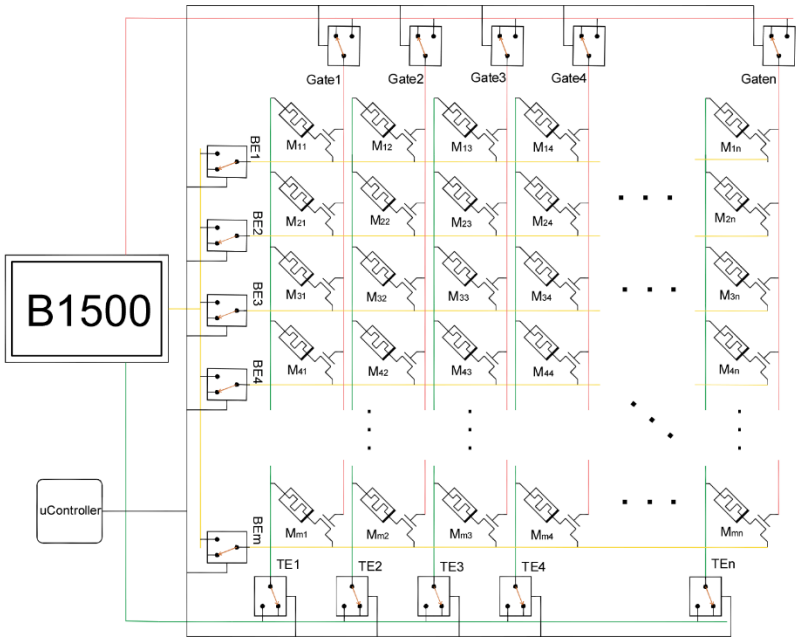


Figure 2.2 : The diagram of switching-matrix measurement system. The system includes micro-controller, relays and other peripheral supporting components.

Fig. 3. 3 shows the finished measurement system, the PCB boards. On it, relays were used as the switches for the reason of avoiding the parasitic parameters, for example, the parasitic capacitance and on-resistance of switches. This system can support 1T1R crossbar array sizes up to 32x32, also it is passive crossbar array compatible.

Connectors on the right side of red box are high speed connectors which uses parallel coaxial cables to connect with a probe card, as the Fig. 2.3b shows a 32x32 probe card which connects with 1T1R crossbar array. Left connectors are designed to act as the input of external driving sources, which can be connected to multiple sources for complicated applications. In the left red box in the Fig. 2.3a are the common terminals which can connect to gate, top electrode, bottom electrode and ground of the crossbar array respectively. They also connect to the external driving sources, such as semiconductor analyzer B1500A. A controlling board was designed under the switching matrix which has a microcontroller and other support circuits to provide the switch of relays between ON and OFF. In default mode, relays are connected to floating pins. The microcontroller receives commands from a computer via a serial port. Controlling program on the computer is an upgrade version of the program used in controlling the B1500A. As it was implied by in the previous discussion, the switching matrix is an extensive part for B1500A.

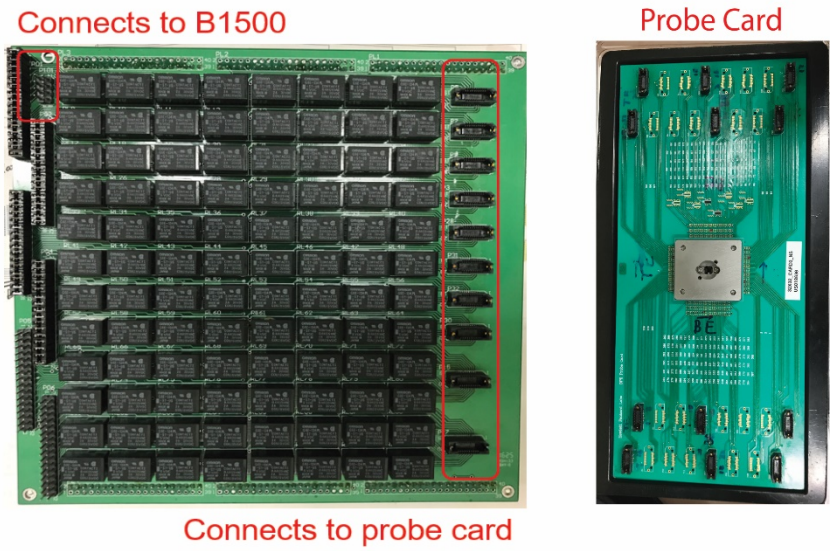


Figure 2.3 : The Switching-Matrix and 1T1R probe card. Each black box on the board represents a relay, through controlling its turn on or off to access any device in the 1T1R array which connects with probe card.

An example interface of the program for controlling the switching-matrix is shown in Fig. 2.4. This program was coded by the C# language. Except the fundamental operations of accessing single device, the biggest change is offering supports to crossbar arrays. It can be used to operate all devices in an array one by one, while collecting operational data to do further analysis. Using buttons in the right panel to choose which device to be accessed, multiple selected devices can be accessed at the chosen sequence. Also, specific pattern can be programmed into an array, as the interface shows in the Fig 2.4, shapes of letter “U” and “M” are going to be programmed.

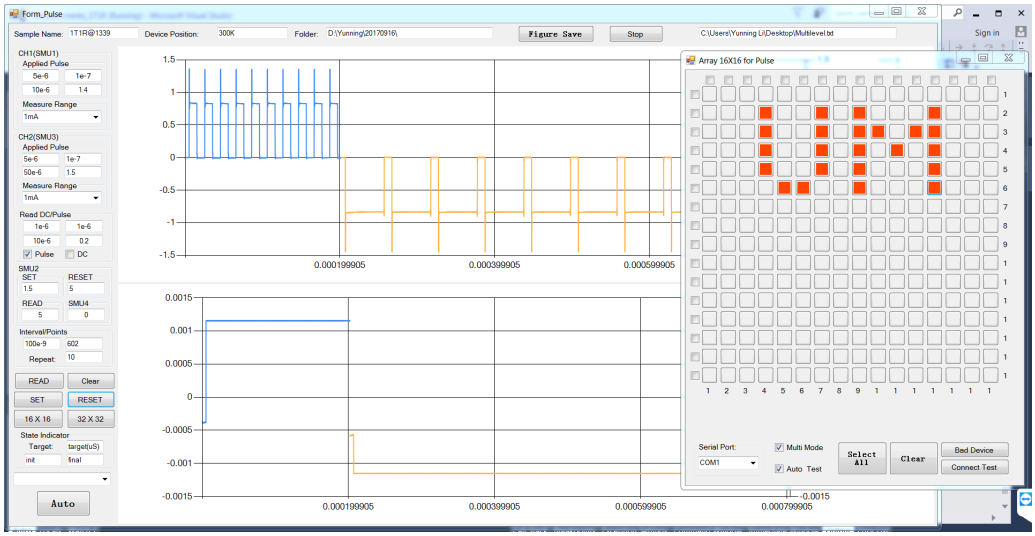


Figure 2.4 : Interface of the controlling program. Using buttons in the right panel to choose which device to be accessed, multiple selected devices can be accessed of the chosen sequence.

Unlike controlling by user interface, only binary states can be used to program. A configuration file can be uploaded to the program with either analog or binary states, meanwhile the program can use a prepared algorithm to tune the state of devices. Here is an example where this system is used to program a specific pattern. In Fig. 2.5a, a target pattern is presented in the binary value and in Fig. 2.5b is shown the written result. In the form b resistance value lower than 5k represents the binary 0, then higher is binary 1.

This experiment was tested under a 16x16 1T1R array, it was processed forming first for all devices then to do the pattern programming. From the result, it demonstrated this system can be used for array level devices successfully.

a.

1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	0	1	1	0	1	0	1	1	1	0	1	1
1	1	0	1	1	0	1	0	0	1	0	0	1	1
1	1	0	1	1	0	1	0	1	0	1	0	1	1
1	1	0	1	1	0	1	0	1	1	1	0	1	1
1	1	1	0	0	1	1	0	1	1	1	0	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1

b.

18.5	16.2	278	15.6	11.8	17.4	10.5	10.9	14.3	19.6	10.3	14.9	12.8	16
11.1	13.4	2.2	19	7.37	2.25	19.7	1.45	12	15.2	12.3	0.79	18.9	14.9
12.4	15.4	1.96	16.9	17.7	2.15	14.4	2.14	0.6	19.9	1.19	1.82	17.9	8.01
17.7	274	1.29	10.1	10.6	1.98	16	1.28	5.79	1.42	13.6	0.65	16.6	17.4
14.8	15.7	2.42	11.1	12.9	1.26	11.8	1.62	125	14.5	13.6	2.16	23.2	14.3
9.34	18.2	20	1.75	1.93	15.8	14.1	1.95	17.3	15.1	14.4	3.18	12.9	20.5
19.4	13.8	17.9	22.2	20.2	13.9	14.3	10.5	16.7	15.6	16.8	11.2	16.4	15.2

Figure 2.5 : Program a specific pattern into a 1T1R array via the switching-matrix. Form b, resistance less than 5k is binary 0, then is 1.

2.3 Dot Product Engine (1T1R Measurement System)

As we all know memristor not only shows the resistance switching in a huge ratio, but also exhibits precision resistance tunability within a specific range, for example, HfO2 device could accurately and stably continuously tune in the resistance range from 1k to 10k Ohm. That gives possibilities of using memristor not only as a random-access memory, but also for any application that depends on analog computation. For example, memristors are in neuromorphic or vector-matrix multiplications. There is already have a lot of publications that have demonstrated small scale applications based on analog behavior of memristor. But experimental results of those publication have been generated using B1500A or other instruments as driving sources.

Even B1500A is arguably the best instrument for semiconductor analysis, and the switching-matrix has expanded the B1500A operation from single device to a crossbar

array and has enabled the ability of using it in many computing scenarios. But channels either for driving and sensing are still limited. Multi-channel driving is a basic feature in lots of analog applications such as image convolution needs as least nigh channels of driving sources. Here we report a new measurement system that can provide up to 128 channels voltage driving and 64 channels current sensing. The multiple channel voltage driving, and current sensing can be implemented on the system simultaneously. The rough diagram of this system is shown in the Fig. 2.6, driving voltages are generated by high resolution digital-analog converters (DACs) and uses trans-impedance amplifiers (TIAs) and analog-digital converters (ADCs) to sense current. Main features of this system are listed as followed:

1. Can provide full program functions of memristors, such as FORMING, READ, SET and RESET operations.
2. Generate multi-channel driving voltage applied on devices and can also sense output current simultaneously.
3. High precision of voltage outputs and current sensing (16 bits DACs and ADCs).
4. Wide voltage output and current sensing range (-10V~10V, 0~2mA).
5. Support array size up to 128x64.
6. Adjustable pulse width (Shortest 180 ns and 4 ns per step).
7. Compatible with passive and 1T1R array.

It can be seen that voltage generated by DAC not only applies on bottom electrodes and gates of 1T1R crossbar array, but also on top electrodes. In addition, the current sensing combination of ADC and TIA also connects with top electrode. This

configuration implies that for different device's operations an alternative electronic circuit part is connected into the crossbar array.

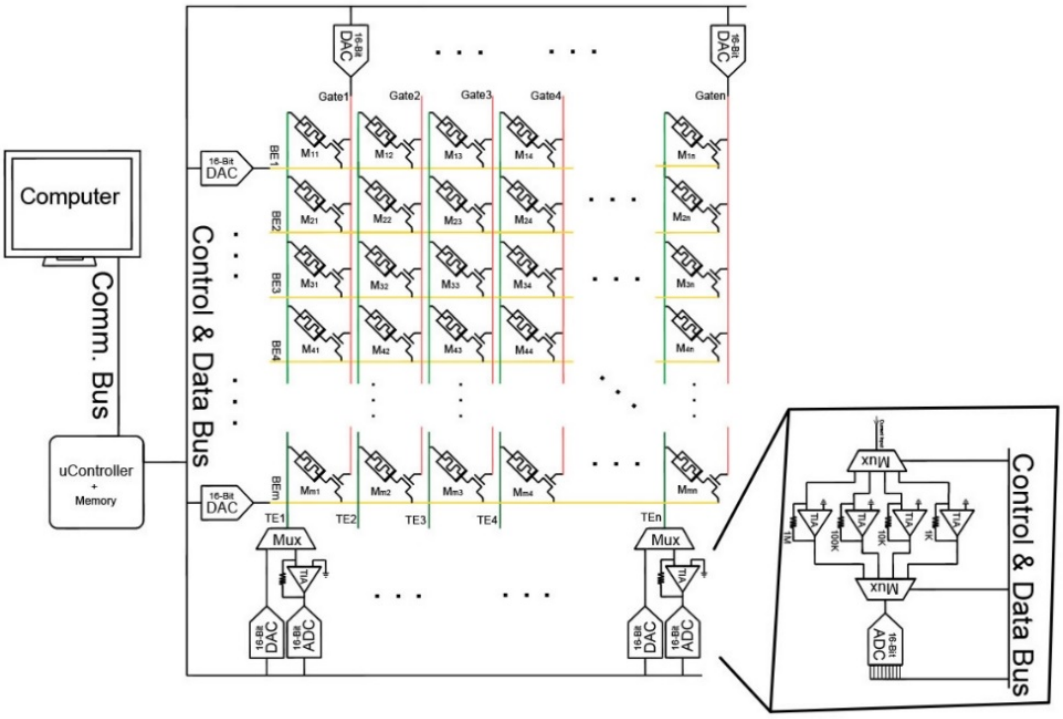


Figure 2.6 : The 1T1R measurement system or DPE system. This system uses high resolution DACs and ADCs voltages generating and current output sensing. TIAs have four configurations to cover different current range, 2mA, 200µA, 20µA, and 2µA.

For example, for the READ and RESET operations, voltage was applied on BE terminal, meanwhile connecting TE terminal to TIA or ground which was controlled by the multiplexer. On the other hand, the SET operation was done by applying voltage on TE and by connecting BE to ground. It is worth mentioning, that each TIA module is a parallel combination of 4 TIAs, as the sub-box shown in Fig. 2.6, with the purpose of covering different current range from µA to mA. Also, the system is controlled by a microcontroller which communicates with a computer through serial port. All programs used in this setup are based on MATLAB, however it is a data analysis friendly platform and boasts a lot of powerful tools. This system is designed for applications based on

memristor array, such as vector-matrix multiplication and can generate abundant number of data.

Fig. 2.7 shows the PCB boards of this system, it consists of 8 column boards (the bigger and left ones) and 8 row boards. All boards are installed on a mother board. Each column board provides 8-channel driving voltage for gates of 1T1R transistors and 8-channel current sense for top electrodes, while each row board can provide driving voltage to 16 channels. On the mother board, a microcontroller is integrated which can randomly configure any channel on boards depending on commands.

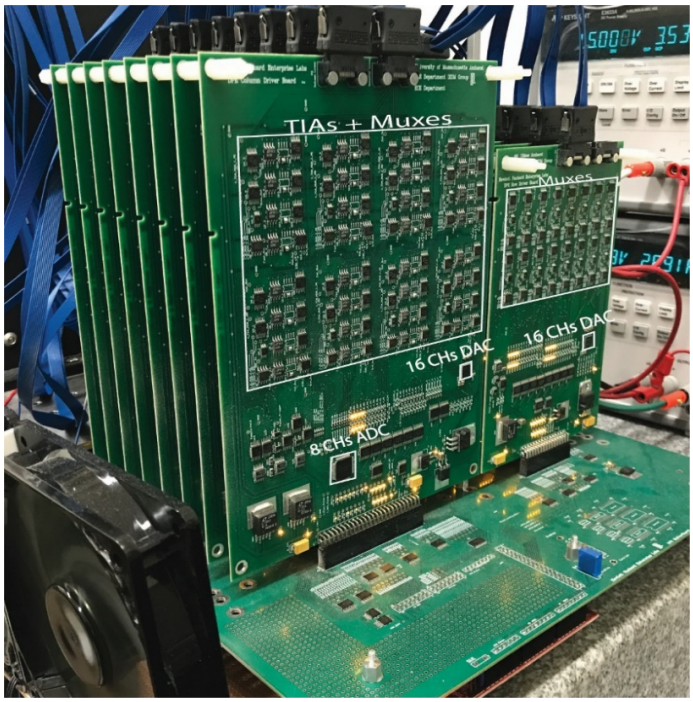


Figure 2.7 : The measurement system – PCB boards. It consists of 8 column boards and 8 row boards. Column board can provide voltage driving for transistor gate and sense current. Row board provides driving voltage only.

As previously mentioned, the system can support an array size up to 128 x 64, Fig. 2.8 is a 128 x 64 1T1R array which connects with a probe card and uses coaxial cables to connect to the system. For this reason, connection pads of device and tips of probe card both are tiny and vulnerable. So, the task of landing the probe card on devices

is a time consuming and painful process. Fortunately, our 1T1R layout has the benefit that bottom electrodes are designed with two pads on both side of the chip. In this case, an indicator circuit was designed, which could light LEDs on it when both sides of tips of probe card could touch with both sides pads of crossbar array.

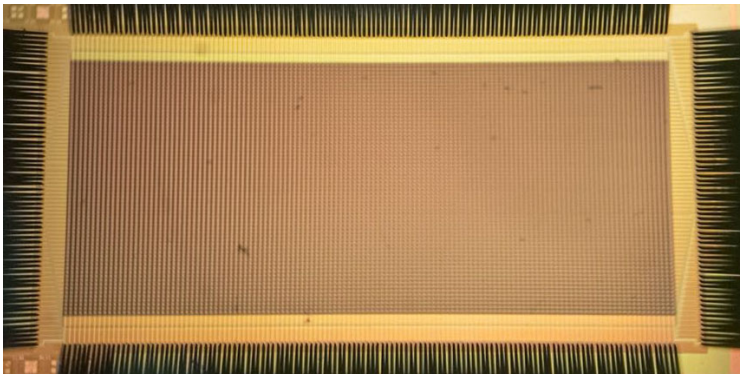


Figure 2.8 : Photograph of 128 x 64 1T1R crossbar array connected with probe card.

The whole setup of this system is showing in Fig. 2.9, it is put on an air-floating table to avoid influence from environmental vibrations. A probe card is installed on the probe station and cables are used to connect it with the measurement system. Two Keysight power supplies provide power for the system, the voltage output of this kind of power supply is a stable one and also with very small ripple voltage. It is an important requirement for the high accuracy of analog to digital or digital to analog conversion.

Fig. 2.10 is a typical TIA diagram, with formula 2.1 showing the simplified way of measuring current through resistor R_x. Where R_{ref} is the reference resistor and R_x is the unknown resistance resistor. V_{out} is the output voltage that connects to an input of ADC, V_{read} provides testing voltage, regular range between -0.2V to 0.2V. Four different resistances of resistor (1kΩ, 10kΩ, 100kΩ, 1MΩ) can be replaced on R_{ref} in order to cover different current range. LTC6268 is a 500MHz ultra-low bias current FET input Op-Amp.

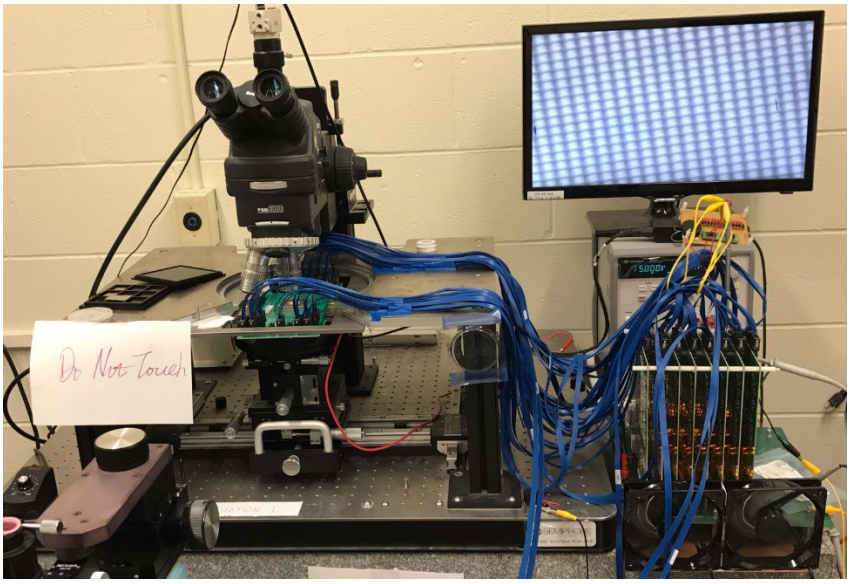


Figure 2.9 : The measurement system or dot product engine. It contains the PCB boards, probe station, and probe card.

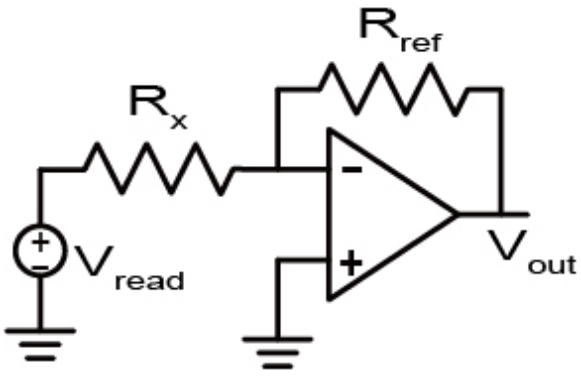


Figure 2.10 : Simplified circuit for resistance measurement

$$V_{out} = -\frac{R_{ref}}{R_x} V_{read} \quad (2.1)$$

The system measurement accuracy has been explored. Here we show experimental results at the resistance of R_{ref} equals to 10k. The under tested resistors are chosen from 500 Ohm to 50k Ohm. There is a parasitic resistance of about 40 Ohm for each channel, which has been confirmed. It comes from the cable resistance and multiplexer on-resistance. Fig. 2.11 shows the distribution of the 10,000 times current

read results but was converted into conductance style plot in the figure. The mean value μ and standard deviation σ has been calculated. It is observed that the resulting conductance are some discrete values. The resolution of DAC and ADC determine this. The plotting data shows that the system is highly accurate. The measurement error is within 1%.

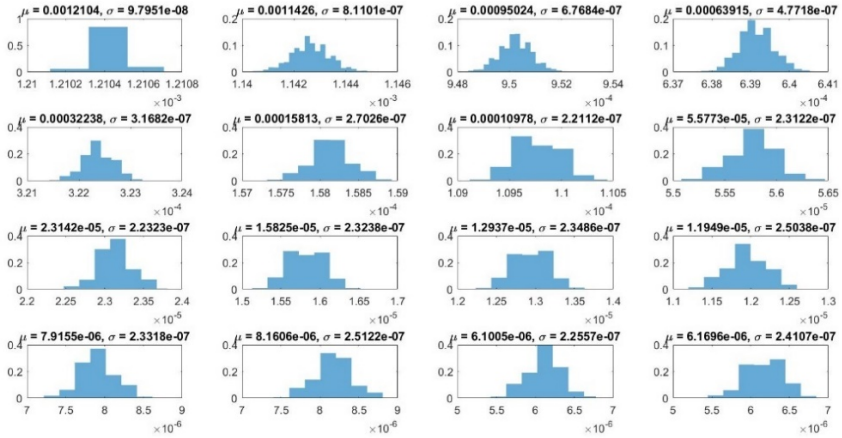


Figure 2.11 : Distribution of measurement of fixed value resistors. The system with high accuracy on current measurement with 1%.

The system is developed for matrix computation, but the computation accuracy is determined by the device's stability, continuous linear conductance and the resolution of driving voltage and reading circuits. An accurate programming algorithm was designed to map arbitrary matrix values appropriately to memristor conductance in a realistic crossbar array. This algorithm uses the close-loop tuning and the adjusting of access transistors. During this process, each tuning cycle is defined as the iterative step of reading and writing. At the beginning, program based on the algorithm reads back a conductance map from a crossbar array; then does the comparison with target conductance value for each device in it to generate a voltage map that indicates what kind of voltage will be applied on the corresponding device. For example, if the read conductance is larger than target value, a reset voltage will be applied. The next step is to

do the writing step and it depends on the decided voltage map. If there is a device that didn't show response for the applied voltage, for the next cycle the operating voltage will increase or decrease a little for SET and RESET respectively. In the SET process, if the device is still without response for stimulus, the gate voltage will be increased. Figure 2.12 exhibits three samples of tuning memristor conductance, second target conductance higher than the real conductance, system used increasing set voltage to gradually tune device to target value; second, the device was reset gradually into the target value; the first one is to exhibit the ability of automatically tuning the conductance, when the occurrence of over tuning program can correct itself spontaneously.

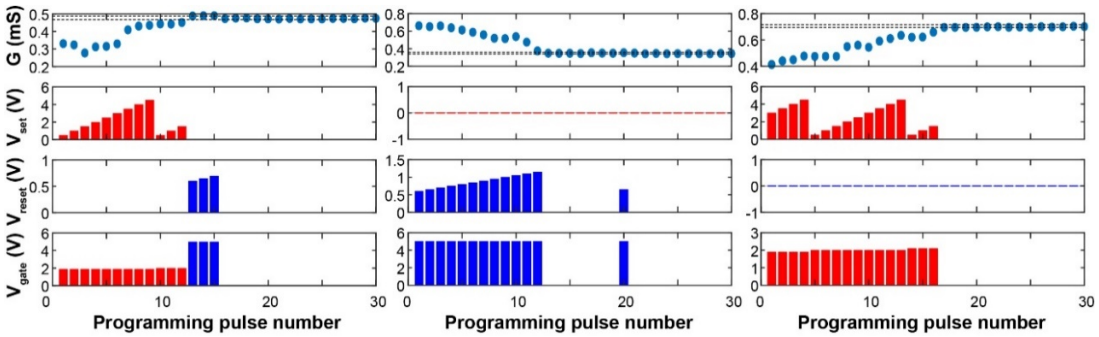


Figure 2.12 : Examples of using feedback-loop tuning conductance of memristors in an 1T1R crossbar array.

2.4 Summary

The unique electronic property of memristors reveals its potential for wide utilization in different applications, such as non-volatile memory, analog computing, and neuromorphic computing. Any application based on memristor devices relies on peripheral circuits or measurement systems. B1500A as a general semiconductor analyzer has been discussed in this chapter, which shows limits when studying devices on array scale. To overcome those disadvantages of using the commercial instruments, two measurement systems were designed and built to perform different tasks. For the

switching-matrix, it works with B1500A, which provides an easy way to do statistics of a mass of devices from crossbar array. This system uses relays as switches, by controlling those relays, it can access any single device in a crossbar array. Another system was designed for parallel computing via using 1T1R crossbar array. Which can provide 128 channels voltage driving and 64 channels current sensing simultaneously. High resolution DACs and ADCs were used in this system; it makes the possibility of demonstrating many applications, for instance, the vector-matrix multiplication or neurons response.

A well-developed measurement system could simplify and shorten test processes and time consumption.

2.5 Reference

1 M. Prezioso, *et al.* Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* 521, 61-64, (2015).

2 D. Marsh, *et al.* Programmable analogue ICs challenge spice–and–breadboard designs. *EDN Europe*, 30-36, (2001).

3 L. Gao, *et al.* Fully parallel write/read in resistive synaptic array for accelerationg on-chip learning. *Nanotechnology* 26, 455204 (2015).

4 M. Hu, *et al.* in *Proceedings of the 53rd annual Design Automation Conference.* 1-6 (2016).

5 M. Hu, *et al.* in *2016 IEEE International Conference on Rebooting Computing (ICRC).* 1-5 (2016).

6 M. Hu, *et al.* in *2012 49th ACM/EDAC/IEEE Design Automation Conference (DAC).* 498-503 (2012).

7 M. Hu, *et at.* The Dot Product Engine: a memristor-based experimental platform for computational acceleration. (2017 in preparation).

8 M. Hu, *et al.* in *Design Automation Conference (DAC), 2016 53nd ACM/EDAC/IEEE.* 1-6 (IEEE).

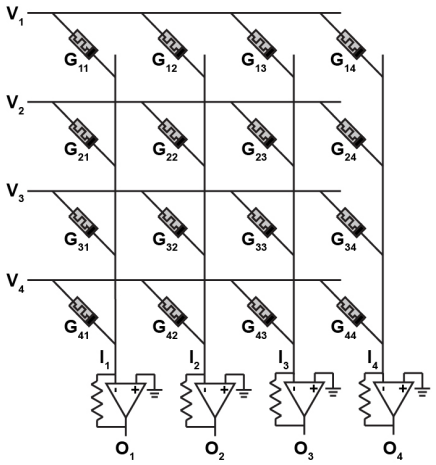
CHAPTER 3

ANALOG COMPUTING AND IMAGE PROCESSING

3.1 Analog VMM introduction

In mathematics, vector-matrix multiplication (VMM) is an operation which expresses the process of producing one matrix from the multiplication of two matrices. It has numerous applications not only in applied mathematics, but also in science and engineering. VMM is also a universal algorithm widely used in computing for a variety of tasks such as image or video signal processing, computer vision, or neuromorphic computing related applications. However, in digital computing system VMM is a tedious and hardware intensive task, as the computation is operating on the clocked hardware in sequential manner, computation complexity grows in $O(n^2)$ and cannot be simply reduced where 'n' is the dimensional size of matrix. To overcome challenges of digital VMM, a memristor crossbar array based analog accelerator for VMM was introduced which can perform in a highly parallel way and consumes very low power. In contrast to using digital binary numbers, an analog accelerator encodes numbers using the continuously-valued circuit voltage and conductance of memristor. Furthermore, the computation complexity can be reduced from $O(n^2)$ to $O(1)$. That predicts the high computing performance when using a large size of memristor array. The main reason for the speed-up is using the intrinsic electrical properties of memristor crossbars when doing VMM on it. As shown in the Figure 3.1, when applying a vector of voltage signals on the rows of a memristor crossbar, multiplication is carried out by each memristor. Each memristor's conductance is multiplied by the voltage drop on it determined by Ohm's Law. According to the Kirchhoff current law (KCL), the current is summed across each

column. The output current reflects the result of multiplications and summations, only takes one constant time step. The current output can be converted into voltage by a trans-impedance amplifier (TIA), while it can either be digitalized by ADCs for usage in digital systems, or forwarded to other analog systems, like crossbar arrays, analog sensors or actuators.



$$\mathbf{I} = \mathbf{V} \cdot \mathbf{G}$$

Example:

$$I_1 = V_1 G_{11} + V_2 G_{21} + V_3 G_{31} + V_4 G_{41}$$

Figure 3.1 : Schematic of the analog VMM operation and mathematics formula. Multiplication takes place in the crossbar array based on the Ohm's law, current as the output is the sum at each column according to the Kirchhoff current law.

In contrast to traditional ASIC matrix multipliers, the vector matrix multiplication and the conductance writing of memristor crossbar takes place and is stored at the same location. This way can help to alleviate the bottleneck issue of communication which existed in von Neumann Machine. There are also other benefits of using memristor crossbar for VMM, among which the most important one is the energy efficiency compared to traditional digital computing.

3.2 Discrete Cosine Transform (DCT) in Imaging Processing

As the involvement of computers in our daily life continues to grow, also does the need for efficient ways of processing and storing large amounts of data. For example,

when a web page is loaded, dozens or perhaps hundreds of images may transmit through the limited network bandwidth. In this case if images can be compressed first, then time consumption on transmission can be tremendously reduced. The JPEG is a widely used form of lossy images in the network, and discrete cosine transform (DCT) is an efficient way to compress lossy images. A DCT is a Fourier-related transform similar to the discrete Fourier transform, but using only real numbers. DCT converts the sampled signal or function from its original domain (time or position) to the frequency domain. As the most important discrete transform, it has been used to perform Fourier analysis in many practical applications including mathematics, signal, and image processing. Formally, the DCT is a linear, invertible function, mathematically it can be expressed as shown in the formula 3.1, 3.2.

$$y(k) = w(k)\sum_{n=1}^{N} x(n)\cos\left(\frac{\pi}{2N}(2n-1)(k-1)\right),\ k = 1, 2, ..., N, \tag{3.1}$$

$$w(k) = \begin{cases} 1/\sqrt{N}, & k = 1, \\ \sqrt{2/N}, & 2 \le k \le N \end{cases} \tag{3.2}$$

Where the x(n) is the input vector signal, other part expresses the DCT matrix. Typically, using DCT in lossy image compression, two times (or two-dimensional) of DCT computation are required applying on the original images. For the first time, the original image was input into the DCT matrix and the output is the image information has been converted into frequency domain and concentrated to the left side (low frequency domain) of the frequency domain map. The second time DCT computation is using the first-time output as the input but rotate 90 degrees, which can get more concentrated frequency information of the image to a corner of frequency domain map. Fig. 3.2 demonstrated the process of two-dimensional image compression, the valuable frequency

information for the image is on a lower frequency corner (Step 2). The remaining part of the frequency map with very small values is less important information. The next step called quantization or compression process, a tradeoff between the image quality and data size, which decides the ratio of low and high frequency information for keeping or discarding respectively. The decompression of an image from the frequency information is a reverse process of DCT, as step 3 and 4 shows in the figure, but using a different computing pattern – Inverse DCT. As a result, the reconstructed images contain some distortion, but the level can be adjusted by the ratio of the compression stage.

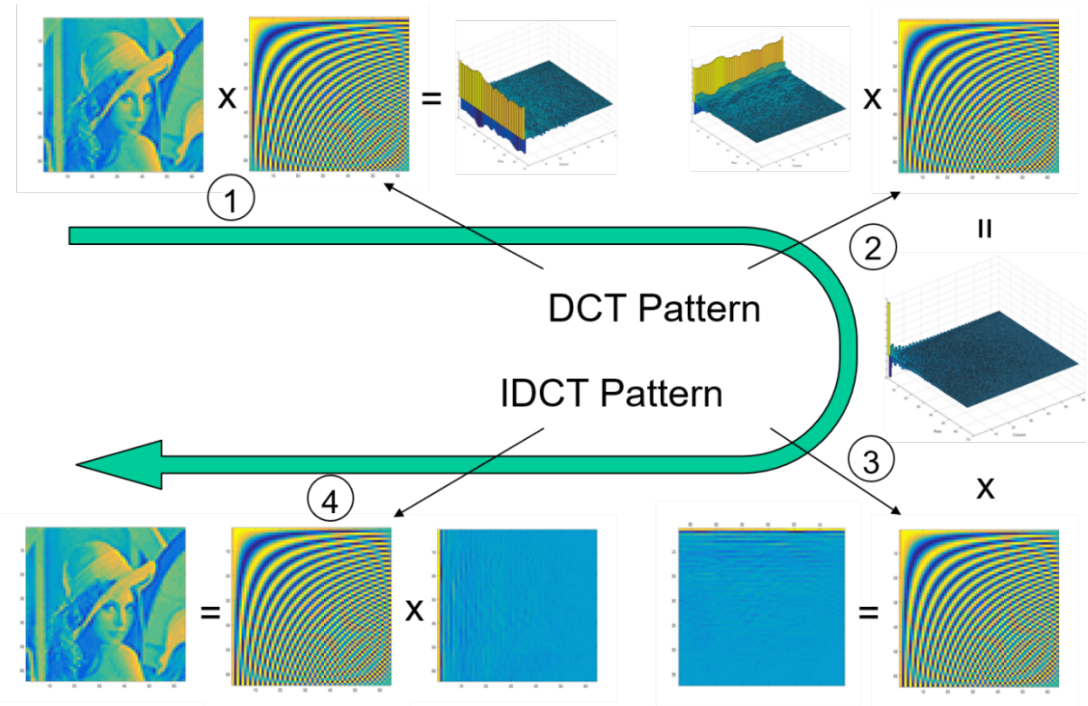


Figure 3.2 : The process of JPEG image compression and decompression. Step 1 and 2 are the processes of 2D DCT for image compression, the second-time input is from the output of first time and is rotated 90 degrees. Step 3 and 4 demonstrate the decompression via using inverse DCT pattern.

3.3 Program DCT pattern on Memristor Crossbar Array

The non-volatile resistance states and reconfigurable memristor crossbars provide a speed and energy efficient way for VMM operation which is widely used in various

computation scenarios. In the following sections of this chapter, we experimentally demonstrate applications of spectrum analysis for image compression and convolutional filtering using memristor crossbar-based computing elements, for sizes up to 128 x 64 cells. The DCT matrix was first configured into the crossbar array, see Fig. 3.3. Because the DCT uses cosine functions, so the value range of this matrix is between -1 to 1. In this case, the main challenge in implementing the DCT in a crossbar is that memristors cannot be used to represent negative values, for device's conductance must be a positive value. To address this issue, a shift relationship is used here to map conductance values to cosine function's value by a linear function as shown in formula 3.3.

$$\mathbf{G}_{\mathrm{dct}} = \mathbf{M}_{\mathrm{dct}} \cdot \beta + \mathbf{J} \cdot \mathrm{m_s} \tag{3.3}$$

$$\beta = \frac{[\mathbf{G}_{\max} - \mathbf{G}_{\min}]}{[\max(\mathbf{M}_{\mathrm{dct}}) - \min(\mathbf{M}_{\mathrm{dct}})]}; \quad \mathrm{m_s} = \mathrm{G}_{\min} - \beta \cdot \min(\mathbf{M}_{\mathrm{dct}}) \tag{3.4}$$

$$\mathbf{OUTPUT} = \frac{\mathbf{G}_{\mathrm{dct}} \cdot \mathbf{X} - \mathbf{J} \cdot \mathrm{m_s} \cdot \mathbf{X}}{\beta} \tag{3.5}$$

Where **J** is a matrix with all ones, and $\beta, \mathrm{m_s}$ are the shift coefficients determined by formula 3.4. For the Fig. 3.3b is the real conductance map read from the upper half of a 128 x 64 crossbar array. Among the 8,192 devices we get 99.8% yield accounting for responsive devices, there were only 18 stuck devices. The histogram of the programming error, defined the difference between the target conductance value and measured value in the crossbar array, is plotted in Fig. 3.4. The writing error followed a normal distribution with a standard deviation of 6 uS. Between the conductance range 100uS – 900uS, 64 levels or 6bits of digital precision was demonstrated, which is sufficient for many applications such as machine learning and analog computing. For typical JPEG image, the value for each pixel is between 0 ~ 255. It cannot be directly provided to the crossbar array and must be converted into voltage that drops in the linear operational range of

memristor between -0.2V to 0.2V. The DCT result can be recovered by using the equation 3.5, where **X** is the input vector voltages. Obviously, the output of crossbar array cannot be directly used as the result of DCT if without post-processing. This process will add to the complexity of circuits design and nullify the advantages of using analog VMM.

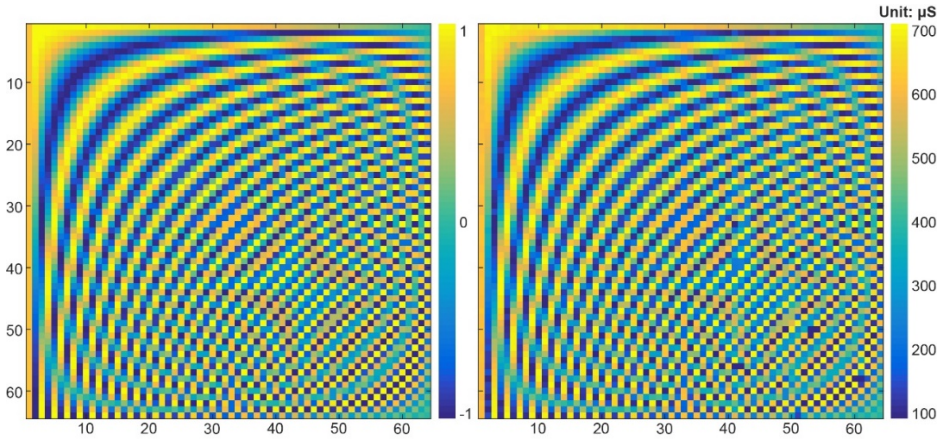


Figure 3.3 : Precise conductance writing in a memristor crossbar array. Left is the target conductance map generated by the MATLAB script. Right is the real conductance reading from a 128 x 64 1T1R crossbar array, only keep the upper half.

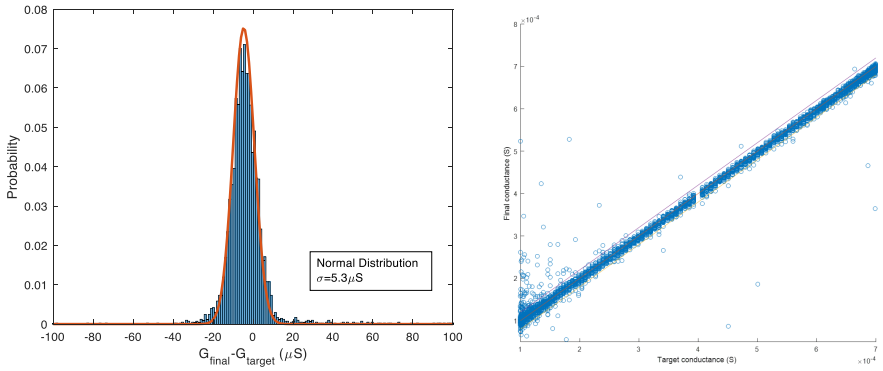


Figure 3.4 : stochastic analysis for writing conductance. Left is the histogram of all devices' error conductance distribution. Right, the comparison between target conductance and the writing results. The lines show the tolerance value during the conductance programming process, typical ± 10uS.

As a spectrum analyzer, the one-dimensional DCT was implemented on the crossbar array, where the input signals are sine waves with different frequencies. Then

the expected output of the crossbar should display different frequency spectrum for input sine wave with various frequencies. As the Fig. 3.5a shows, the input signals are discrete sine waves in voltage and without adding any DC offset, three frequencies were chosen. Fig. 3.5b and 3.5c exhibit the experimental outputs and mathematic simulation outputs. Outputs from two sources are showing good agreement.

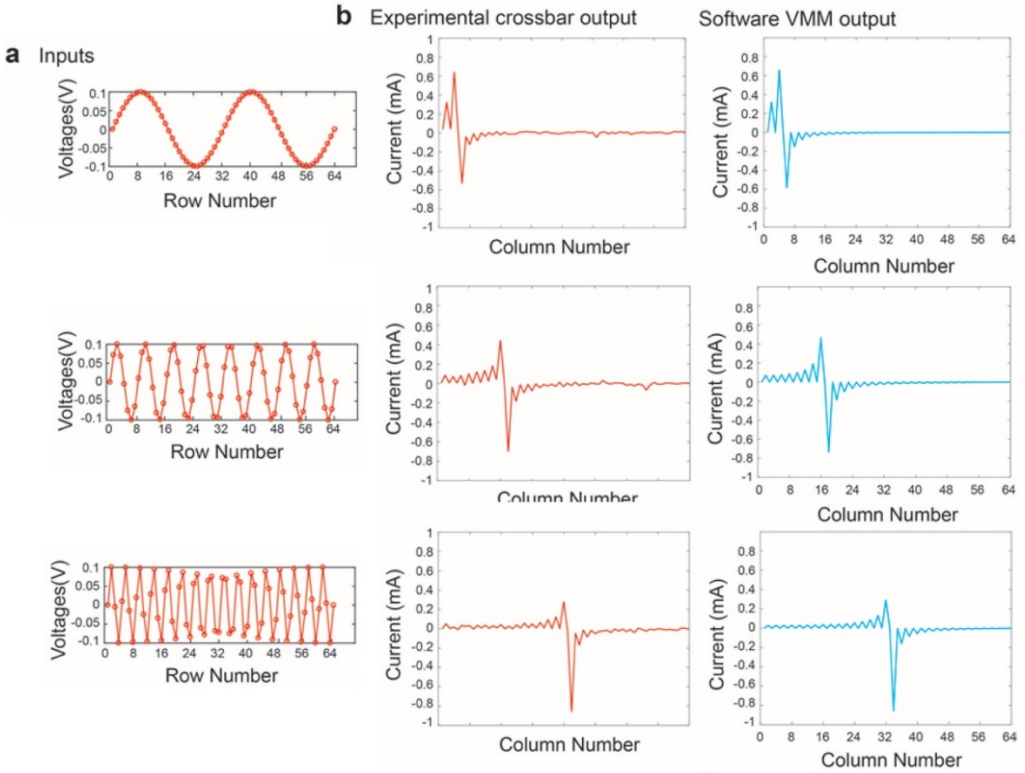
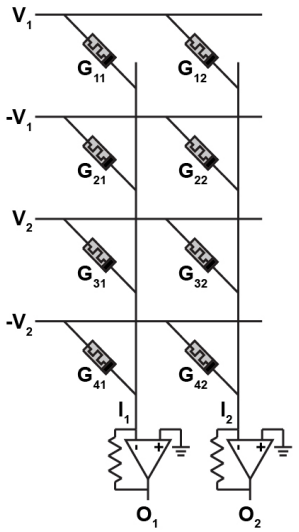


Figure 3.5 : Experimental spectrum analyzer realization on memristor crossbar for sine wave. Input sine wave with different frequencies and experimental and software simulation shows compatible results.

To overcome the extra post-processing for directly mapping DCT pattern into crossbar, another approach was employed which uses differential conductance of two memristors to represent one computing element in the matrix. The input voltage signals also expand to using two voltages at same amplitude but of opposite polarities to express one input. The novel approach can be explained using the Fig. 3.6 and formula in it. For example, where ($G_{11} - G_{21}$) is a differential pair. This way can widen the computing

conductance range and represent negative value easily. For example, the common programmed conductance range for memristor is from 100uS to 700uS, only 600uS is valuable using in computing, but for differential pair it could expand to 1200uS from -600uS to 600uS.



$$I_1 = V_1G_{11} - V_1G_{21} + V_2G_{31} - V_2G_{41}$$
$$= V_1(G_{11} - G_{21}) + V_2(G_{31} - G_{41})$$

Figure 3.6 : Memristor crossbar array based differential mode analog VMM. This way can improve the conductance available range and using to represent negative value.

Table 3.1 : The comparison between normal and differential mode for VMM. A mathematic evident of using differential mode reduce the complexity of post process.

	Normal Mode	**Differential Mode**
Conductance Map	$\mathbf{G}_{\mathrm{dct}} = \mathbf{M}_{\mathrm{dct}} \cdot \beta + \mathbf{J} \cdot \mathrm{m_s}$	$\mathbf{G}_{\mathrm{dct1}} = \frac{1}{2}\mathbf{M}_{\mathrm{dct}} \cdot \beta + \mathbf{J} \cdot \mathrm{m_s}$ $\mathbf{G}_{\mathrm{dct2}} = -\frac{1}{2}\mathbf{M}_{\mathrm{dct}} \cdot \beta + \mathbf{J} \cdot \mathrm{m_s}$ $\mathbf{G}_{\mathrm{dct}} = \mathbf{G}_{\mathrm{dct1}} - \mathbf{G}_{\mathrm{dct2}}$ $= \beta \cdot \mathbf{M}_{\mathrm{dct}}$
Computation	$\mathbf{DPE}_{\mathrm{out}} = \mathbf{X} \cdot \mathbf{G}_{\mathrm{dct}}$ $= \beta \cdot \mathbf{X} \cdot \mathbf{G}_{\mathrm{dct}} + \mathbf{J} \cdot \mathbf{X} \cdot \mathrm{m_s}$	$\mathbf{DPE}_{\mathrm{out}} = \mathbf{X} \cdot \mathbf{G}_{\mathrm{dct}}$ $= \beta \cdot \mathbf{X} \cdot \mathbf{M}_{\mathrm{dct}}$
Post-Process	$\mathbf{OUTPUT} = \frac{\mathbf{DPE}_{\mathrm{out}} - \mathbf{J} \cdot \mathbf{X} \cdot \mathrm{m_s}}{\beta}$	$\mathbf{OUTPUT} = \frac{\mathbf{DPE}_{\mathrm{out}}}{\beta}$

There is another benefit of using differential mode which can dramatically reduce the complexity of post-processing for the directly output from the crossbar array, a

competition between normal and differential conductance map mode is listed in the followed table. Obviously, the output of differential mode and the software output is just a scaling relation without the extra summation part, which may only need simple process that could act as the input for other functional circuits.

3.4 Image Compression using memristive DCT pattern

As previous section mentioned, the memristor crossbar array can be used to do image compression. The image used in this experiment is the famous portrait of Lenna, for which the image pixel intensities were converted to voltage signals then input to the memristor crossbar row by row, as the Fig. 3.7 shows.

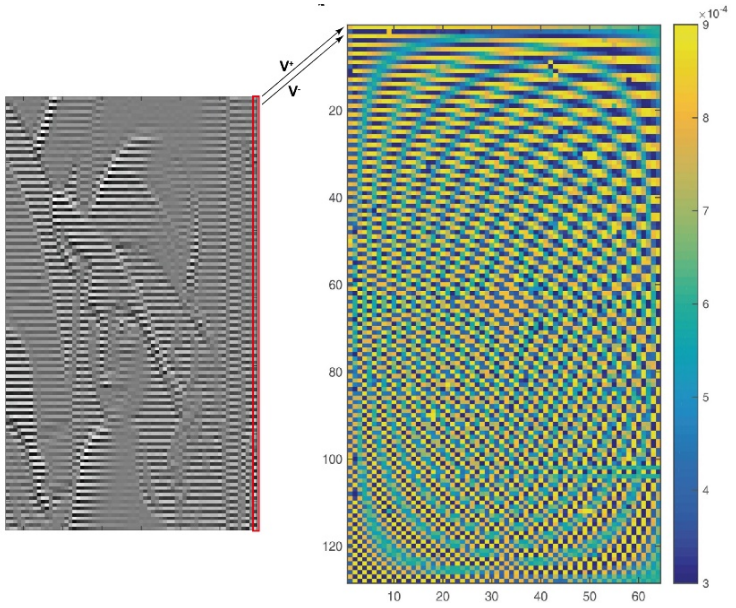


Figure 3.7 : Experimental demonstration using differential conductance pairs for image compression. The image was converted to voltages that were applied to the rows of right pattern. Neighboring rows have a voltage pair with the same amplitude but different sign. Conductance map two elements from two rows represent one value in the cosine DTC pattern.

To achieve a higher density of frequency information, the second DCT is needed by using first time DCT outputs then input the pattern column by column. If the input

image size is larger than the computing crossbar size, the image can be broken into few blocks, each block is 64 x 64 in size. Input blocks to the crossbar is in the sequence of left to right, top to bottom, as shown in the Fig. 3.8. Then tiled output for each block together to achieve the complete information about the image.

For a JPEG image contains red, green, and blue colors information and each color is stored in the same layer. This provides a convenient way to process the image, as the separation of large image to different blocks, the multilayer image can be input to the computed VMM pattern one layer after one. Frequency information was also separated into block by block and layer by layer, and finally recombined together.

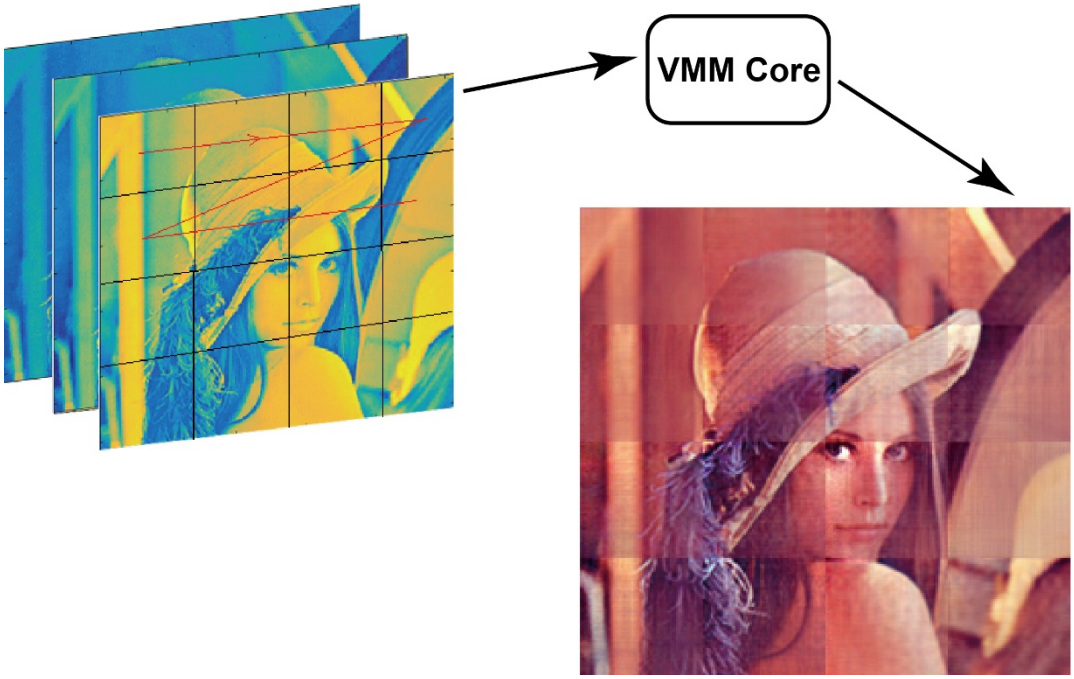


Figure 3.8 : Experimental 2D DCT demonstration for large size of image input. The original input image with size larger than the analog VMM size, which was separated into different blocks and layers. Each block was input the VMM sequentially.

The experimental image compression process is illustrated in Fig. 3.2, the first two steps were realized by the memristor based DCT The DCT pattern generates the

concentrated frequency domain information. Then the last two steps or the reconstructed process of image using MATLAB generated 2D inverse DCT function and the frequency domain information to complete. The recovered Lenna portrait is shown in the Fig 3.8., which is a 256 x 256 pixels image divided into 3 layers and each layer divided into 4 x 4 blocks.

The results are compared with those using MATLAB to do 2D DCT and compression, and the crossbar array outputs by the same data ratio. Details are shown in the Fig. 3.9. Different compression ratios ranging from 20:1 to 2:1 were analyzed and compared. Even with 5% of the original information, the reconstructed image still showing a reasonable image.

3.5 Image Convolution realized on memristor crossbar array

Another popular algorithm often used for image processing is convolution which was also experimentally demonstrated in this work. Convolution serves as the basic computational primitive for various associative computing tasks ranging from edge detection to image matching. In the convolution operations, a vector multiplication is computed between a computing kernel and the corresponding pixels of the input image for each position of the kernel. The total number of convolution operations to generate the final output is large as the kernel sweep across the entire image, each pixel in the converted map corresponds to one VMM operation. The convolution kernel repeats this process by sliding it over the original image until the output matrix is generated. For example, an image convolution process is demonstrated in the Fig. 3.10, there is a Laplacian transform to get edge information of an input image.

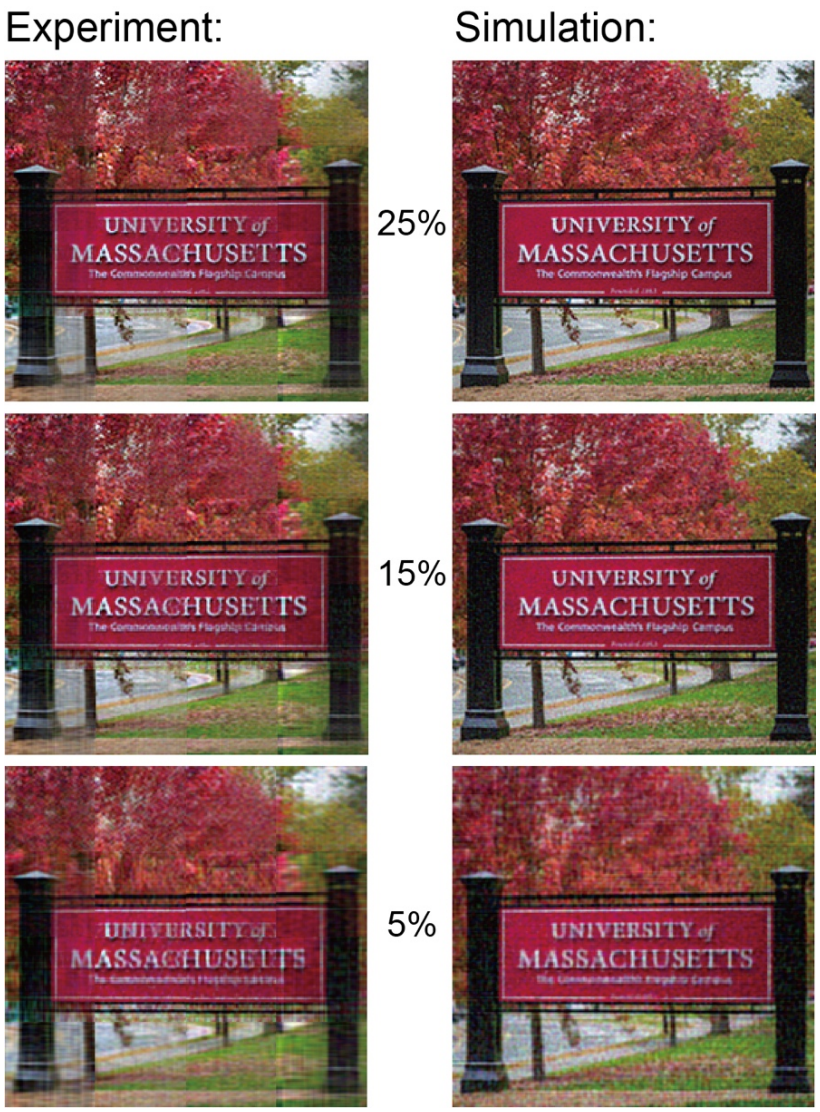


Figure 3.9 : Image compressed with different compression ratios. Top panels are results from software VMM and bottom panels are results from the memristor crossbar.

In this experiment, seven different convolutional kernels are employed. They include Gaussian, Average to eliminate noise in images, LOG and Sobel to extract the edge or mimic the motion blur effect. Differential pair also used here to expand the value represents in negative values. Those pairs were arranged in the neighboring columns rather than rows. As the Fig 3.11 shows, different target convolution kernels were generated, and the same pattern was written into a crossbar array for several times. Then

devices in a region with the lowest conductance error was chosen (in the red box) for the next step measurements.

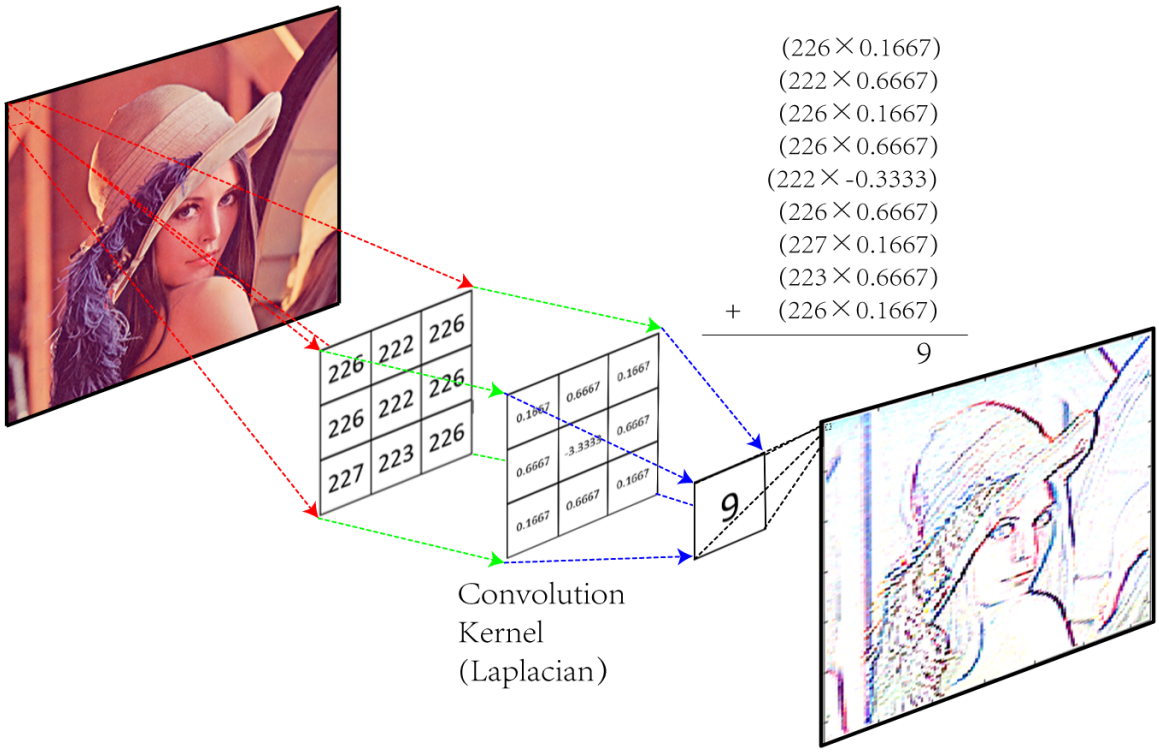


Figure 3.10 : Graphical representation of 2D spatial convolution with 3 x 3 kernel matrix.

The input image also uses Lenna portrait in an array size 128 x 128 and extra artificial Gaussian white noise added. The computing process is similar as the image compression realized on DCT pattern. First, converter value of each pixel in image to voltage pair have same amplitude but different polarity. Then each time in sequence choose 3 x 3 pixels of positive voltage values input to the first column of convolution kernels. This becomes the input of negative voltages to second column of kernels. The experimental results are presented in Fig. 3.12, showing the abilities of smoothing images and extracting the edges out of the images using memristor crossbars. The edge detection is an important step as one convolution layer in convolutional neural network.

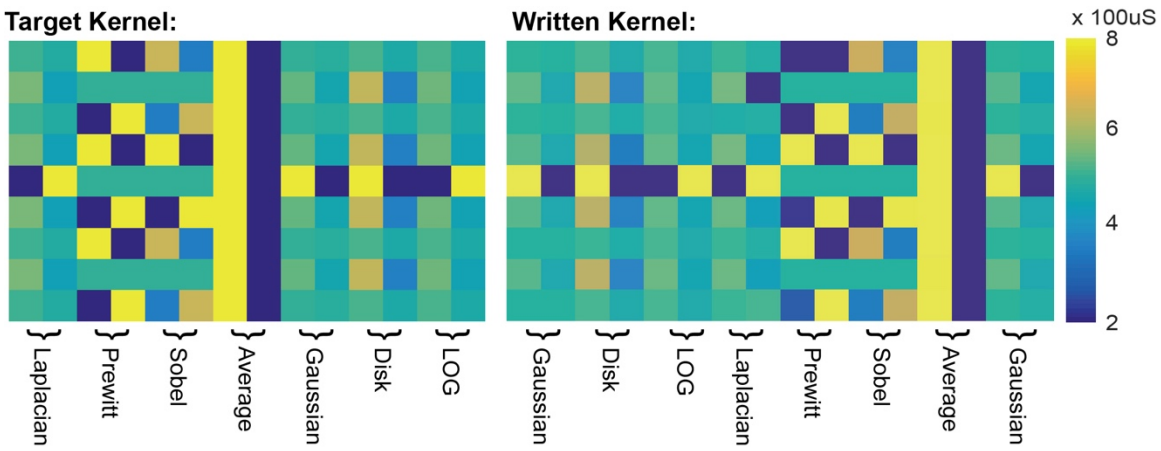


Figure 3.11 : Different convolution kernels using differential mode. Kernels includes Laplacian, Prewitt, Sobel, Average, Gaussian, Disk, and Laplacian of Gaussian. Those kernels were written into the memristor crossbar multiple times, choose the best devices for computation.

3.6 Power Efficiency of Analog VMM

The key advantages of memristor crossbar based VMM are the re-programmable conductance of memristors and the reasonable accuracy and precision within the analog computation, and the high efficiency on speed and energy consumption. For the simplified multiplication based on the 128 x 64 crossbar array, the performance gives 1.6 TOPS (Tera Operations Per Second) when the clock period is 10 ns A circuit-level simulation was designed to calculate the power consumption on image compression based on experimental parameters, including devices' conductance, wire resistance and input images.

It was found that the average power consumed on the crossbar is about 13.7 mW, and the efficiency is about 119.7 TOPS/W. Compared with the highly optimized digital application specific integrated circuit (ASIC) system in edge technology for matrix multiplication, which has the energy efficiency of 7.02 TOPS/W. The analog VMM is 17x more energy efficient than the ASIC solution. Additionally, memristor crossbars

have nearly zero power consumption in their idle state, as the non-volatile properties of memristors, while typical graphic process unit (GPU) still consume about tens Watts when idle. Those features make the analog VMM system is an ideal solution for analog computation.

Figure 3.12 : Experimental convolution results from different convolution kernels. The R, G, B layers of input image were processed separately.

3.6 Summary

In summary, experimental demonstrations for the analog vector or matrix multiplication utilizing memristor crossbars with over 8,000 devices, about 64 level precision and 99.8% device yield have been described in above sections. Device linearity and wire resistance have also been discussed. The most important achievement is the successfully implemented image compression and convolution on the memristor based VMM system. Although, many circuit issues in the analog domain are not trivial, but the system shows reasonable precision, high energy efficiency and computing performance.

3.7 References

1 D. B. Strulov, *et al.* The missing memristor found. *Nature* 453, 80-83 (2008).

2 J. J. Yang, *et al.* Memristive devices for computing. *Nature Nanotechnol* 8, 13-24 (2013).

3 B. Li, *et al.* Exploring the precision for RRAM-based Analog Approximate Computing. *IEEE Design & Test* 33, 51-58 (2016).

4 L. Xia, *et al.* Technological Exploration of RRAM Crossbar Array for Matrix-Vector Multiplication. *Journal of Computer Science and Technology* 31, 3-19 (2016).

5 G. Peng, *et al.* in *The 20th Asia and South Pacific Design Automation Conference.* 106-111, (2015).

6 S. Yu, *et al.* in *Int El devices Meet.* 416-419 (2016).

7 P. M. Sheridan, *et al.* Sparse coding with memristor networks. *Nat Nanotechnol* 12, 784-789, (2017).

8 L. Gao, *et al.* Fully parallel write/read in resistive synaptic array for accelerationg on-chip learning. *Nanotechnology* 26, 455204 (2015).

9 M. Hu, *et al.* in *Proceedings of the 53rd annual Design Automation Conference.* 1-6 (2016).

10 M. Hu, *et al.* in *2016 IEEE International Conference on Rebooting Computing (ICRC).* 1-5 (2016).

11 Y. LeCun, *et al.* Deep Learning. *Nature* 521, 436-444 (2015).

12 M. Hu, *et al.* in *2012 49th ACM/EDAC/IEEE Design Automation Conference (DAC).* 498-503 (2012).

13 S. Park, *et al.* Electronic system with memristive synapses for pattern recognition. *Sci Rep* 5, 10123 (2015).

14 S. Choi, *et al.* Experimental Demonstration of Feature Extraction and Dimensionality Reduction Using Memristor Networks. *Nano Lett 17*, 3113-3118 (2017).

15 C. Li, *et al.* Three-dimensional crossbar arrays of self-rectifying Si/SiO2/Si memristors. *Nat Commun* 8, 15666 (2017).

16 M. Hu, *et at.* The Dot Product Engine: a memristor-based experimental platform for computational acceleration. (2017 in preparation).

17 B. Chakrabarti, *et al.* A multiply-add engine with monolithically integrated 3D memristor crossbar/CMOS hybrid circuit. *Sci Rep* 7, 42429 (2017).

18 H. Jiang, *et al.* Sub-10nm Ta Channel Responsible for Superior Performance of a HfO2 Memristor. *Sci Rep* 6, 28525 (2016)

CHAPTER 4

FPAA

4.1 Introduction of FPAA

In the big data and IoT (internet of things) era, energy efficiency and processing throughput has become increasingly important for complicated computing tasks such as classification and recognition while processing accuracy is less critical. Information might be processed in a much faster and energy efficient manner if they were represented by analog signals, especially the huge amount of raw data generated every second from ubiquitous sensors at the edge of the cloud. Even though the real world is essentially analog, ironically, signals are usually digitized and processed in digital domain. This on the one hand is due to the reduced accuracy of analog circuits, while on the other hand is because of the lack of reconfigurable and scalable platforms for analog circuits, like the field-programmable gate arrays (FPGAs) for digital circuits.

Field-programmable analog arrays (FPAAs) were conceptualized in late 1980s and commercialized in 1996. Such an analog platform could be eventually used as a general-purpose circuit that can be reconfigured directly for different applications or as a tool for prototyping analog designs. The general form of FPAA is a monolithic collection of configurable analog blocks (CABs), a user-controllable routing network (switching matrix) used for reconfiguring the connections of the building blocks, and a collection of memory elements used to define both the function and structure (configuration memory). The CABs may have different elements, such as transistors, programmable resistors/capacitors, op-amps, VMMs etc. However, early CABs lacked basic compact reconfigurable elements such as programmable analog resistors and complicated

solutions were used accordingly which were very costly in terms of chip area, design complexity, noise level and power consumption. The early switching matrices in those FPAAs are CMOS-centric, typically employing pass transistors or CMOS transmission gates, which are volatile and thus necessitate the configuration memory elements in the early FPAAs to store the configuration information. With only a handful function units, these FPAAs resemble the PLD (programmable logic device) stage of FPGAs in the 1980s. The adoption of floating-gate (FG) transistors into the FPAAs was a great step forward towards a large scale FPAA. In these FPAAs, FG transistors were used as both the switching matrices and programmable resistors, leading to a more compact circuit that has more function units yet use less energy. Non-volatility of the FG transistor based switching matrices obviates the need of the memory elements for configuration storage. FG based FPAAs represent the state-of-the-art, nevertheless, they still suffer from a number of issues. Programming a FG transistor in the switching matrices may take up to a few minutes because the device needs to be programmed to the extreme which is out of its normal application regime (e.g in FLASH memory), slowing the operation. In addition, great linearity is required for both switching matrices and programmable resistors so that the resistance does not vary with voltages, which is challenging for FG transistors and thus affects the accuracy. The high operating voltage for generating hot-electrons and poor scalability hinders the energy efficient operation at ultrahigh-density integration. Furthermore, most of the chip area is still occupied by routing networks and it is difficult if possible at all to build a 3D architecture with the routing network layer sitting on top of the CABs. The resistance of the FG transistor switches in their ON state is huge (about 10KΩ) even for large transistors (e.g. 0.5μm technology node). Adding FG

switches in the signal path increases parasitic capacitance and resistance to the path, which will reduce the bandwidth of the system.

4.2 Memristive FPAA

The recent progress on memristive device technology has provided better solutions for large-scale, sophisticated, high-speed, low-energy FPAAs. The key components in the FPAAs, include efficient switching matrices, programmable resistors and vector-matrix multiplier.

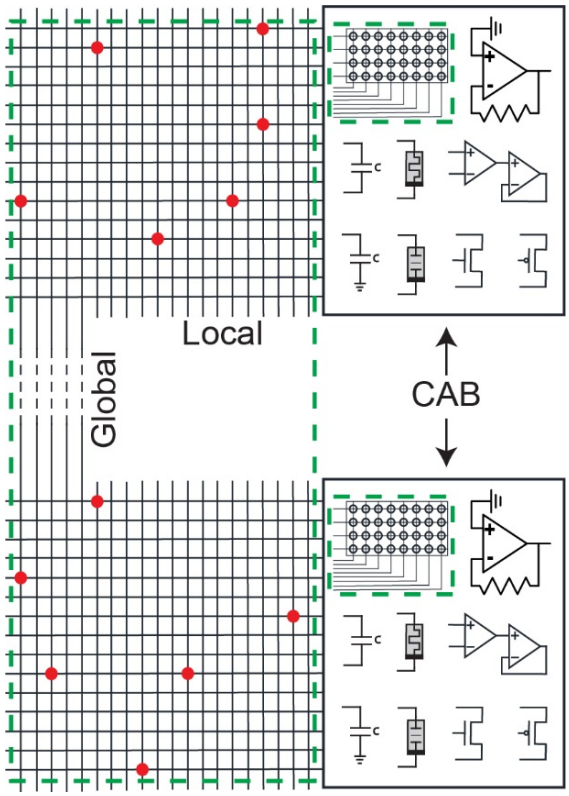


Figure 4.1 :The memFPAA based on 1T1R crossbar array. A memFPAA architecture using a memristor crossbar array as routing network for the global and local networks. A common CAB may consist of transistors, capacitors, TIAs, memristor, mem-capacitors, op-amps as well as analog VMMs based on memristor subarrays of the memFPAA.

All these components can be naturally achieved with memristive devices as demonstrated in this study. Various analog computing tasks, including band pass filters,

audio equalizer and acoustic mixed frequency classifier, have been experimentally implemented in the memristive FPAA (memFPAA) constructed for the first time in this study. As shown in Fig. 4.1, the memFPAA consists of reconfigurable routing network with CABs, in which the routing network is realized by using memristor crossbars. Depending on the granularity level, CABs may be composed of different components, for example, transistors, memristors, mem-capacitors, capacitors, op-amps, as well as VMM etc. Memristors are used as switches on the fabric of the memFPAA. A memristors in its low resistance state (LRS) works as an ON-state switch, which electrically bridges components. On the contrary, the high resistance state (HRS) or the OFF switching memristor shuts the connection so no signal can pass through it. For demonstration purpose, memristive array with 1T1R structure is used in our memFPAA design, comprising a crossbar of integrated hafnium oxide ($Ta/HfO_2/Pd$) memristors built on the drain terminals of n-type enhancement mode metal-oxide-semiconductor (MOS) transistors. But in principle, the access transistor in the 1T1R cell can be replaced by a selector to form a so-called 1S1R cell. The memristors have sufficient ON/OFF ratio to meet the roles of switches in the FPAA fabric. The transistors suppress the sneak path currents during memristor programming, which also enable rapid tuning of conductance of memristors by adjusting the gate voltage to impose current compliance for memristor tuning. More importantly, to fulfil the role of a continuously programmable resistor, each individual memristor exhibits smooth programmability and I-V linearity in the conductance range from 50μS to 1000μS. Such memristors, in the crossbar array, naturally implement the VMM due to Ohm's Law and Kirchhoff's Current Law, featuring low-power and high-density compared to the digital counterpart.

Details of the memristor based FPAA measurement setup are as follows, it is a combination of a 32×32 1T1R crossbar array which connects to external circuit by a probe card and circuits were built on a bread board, which contains op-amps, TIAs and capacitors. The external circuit was powered by a dual-channel voltage power supply. The row and column boards were used to control access devices in 1T1R array and perform memristor conductance tuning. The programmed 1T1R array was switched in to the external circuits. The input signal was produced by a Keysight 33220a waveform generator, while the output was analyzed by a Keysight 3104T oscilloscope. MATLAB scripts were used for automatic data collection and memristor programming.

4.3 First-Order Pass Filter on FPAA

To illustrate the configurability of the memristor based FPAA, we show continuous tuning of the cutoff frequency of the first-order low-pass and high-pass filters, mathematically defined by $|A| = \frac{1}{\sqrt{1+(f/f_0)^2}}$ and $|A| = \frac{1}{\sqrt{1+(f_0/f)^2}}$, respectively, where A is the amplitude gain of output/input, f is the input frequency, and $f_0 = 1/2\pi RC$ is the -3dB frequency point of the filter. A memFPAA based low-pass filter that transmits lower frequencies well but attenuates higher frequencies is depicted in Fig. 4.2a-b, employing a memristor as a reconfigurable analog resistor in series with the load, and a capacitor in parallel with the load. The reactance of the capacitor blocks low-frequency signals. The reactance drops with the increment of frequency which effectively suppresses the high frequency signal. Different from conventional low-pass filter, the programmability of the memristor yields continuous tuning of the cutoff frequency of the filter, as shown in Fig. 4.3a.

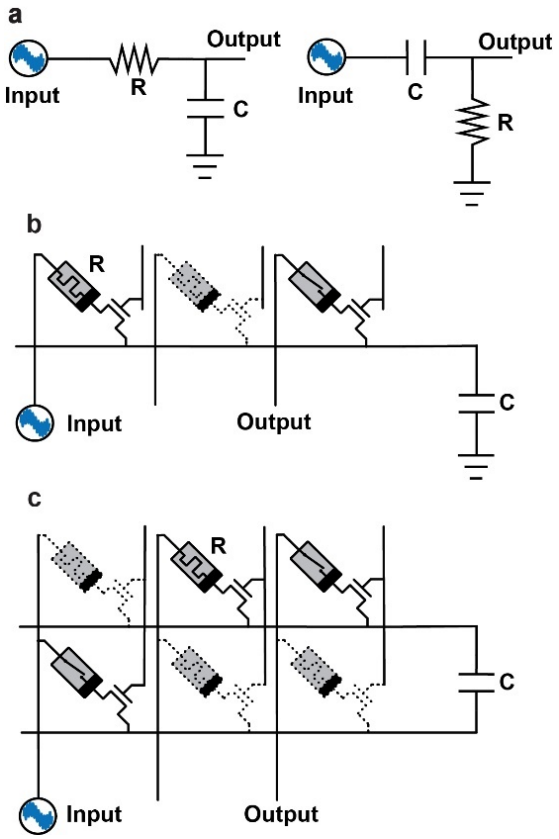


Figure 4.2 : The memFPAA based first-order low/high-pass filters. a. The schematic of a first-order low pass and high pass filter with a resistor in series a capacitor b. The 1T1R array configuration for realizing the first-order low pass filter. The solid line memristor symbol with a switch in side represents an on-state switch, while that without a switch corresponds to the resistance of the filter which shows continuously tunable conductance. The dashed line symbol corresponds to memristor in its OFF state. c. The schematic of a first order high-pass filter, with swapped resistance and capacitance compared to that of b.

For demonstration purpose, the memristor was programmed in the range from 1kΩ to 20kΩ with three different capacitances (10nf, 100nf and 470nf) to cover the whole audio frequency range (20Hz ~ 20kHz in three orders). Similarly, the high-pass filter with continuously tunable cutoff frequency could be constructed by swapping the resistance and capacitance. The observed response spectrum is shown in Fig. 4.3b, which is opposite to that of Fig. 4.3a. A band pass filter could be realized by connecting a high pass filter with a low pass filter in series.

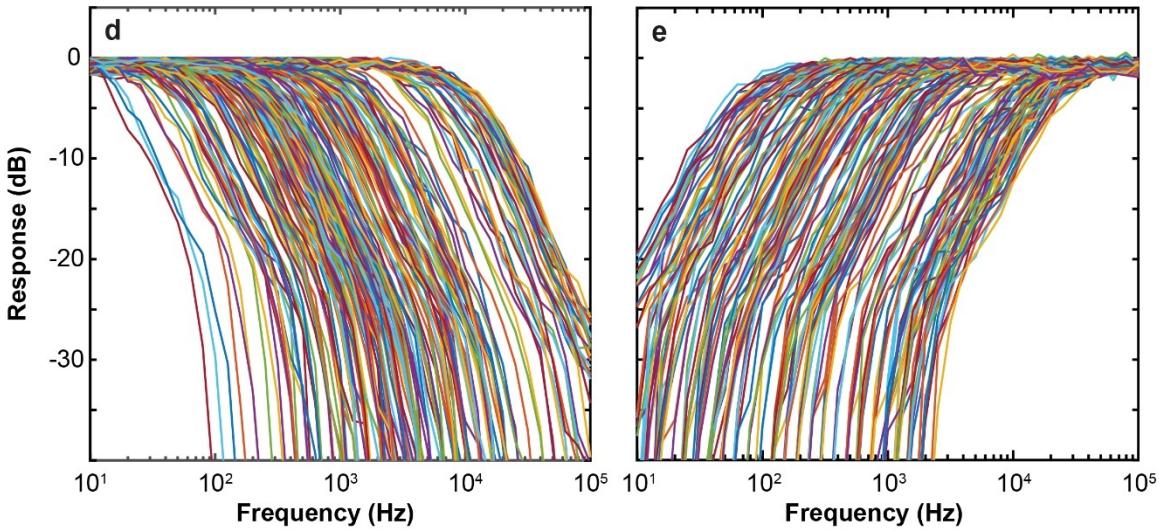


Figure 4.3 : The frequency response of first-order pass filter on the memFPAA. a. The frequency responses of low pass filter with different memristor resistance, covering the full audio frequency range from 20Hz to 20kHz. b. The frequency responses of high pass filter with different memristor resistance.

4.4 Audio equalizer on memristive FPAA

Based on the band pass filters, we further constructed a memFPAA audio equalizer that adjusts the balance of different frequency components of an audio signal. Such equalizer is of wide applications in consumer electronics and telecommunications, either as an independent device in acoustic applications or as a medium-level circuit of a complicated audio system (e.g. pre-processing for speech recognition). A commonly used equalizer architecture in audio processing consists of parallel band pass filters, buffers, and a weight network by VMM, which can be realized using memristors as shown in Fig. 4.4. The band pass filters pick specific frequency bands of the input signal and output to buffers. The weight network modulates the amplitude of each frequency band component and sums them up to produce a signal with adjusted balance. The memFPAA audio equalizer was experimentally built with a 32×32 1T1R crossbar array, external capacitors, and op-amps as shown in Fig. 3b.

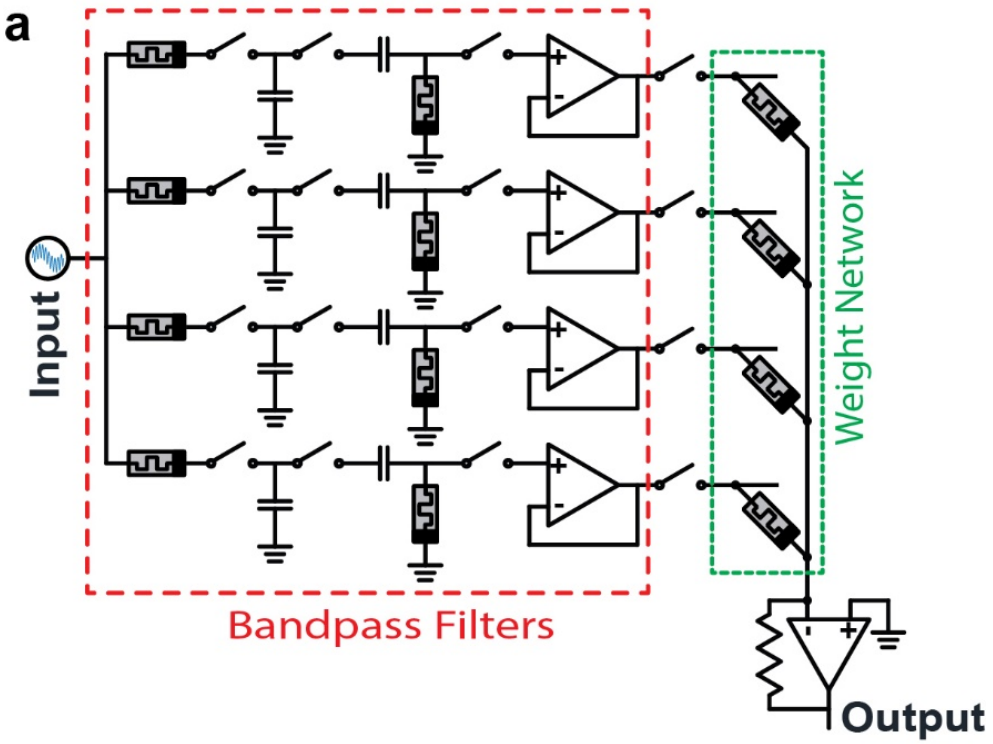


Figure 4.4 : MemFPAA audio equalizer. The schematic a commonly used audio equalizer consisting of a bank of cascaded high and low pass filters followed by a VMM.

As shown in the conductance map, a routing network in an array of size 24 rows by 15 columns was used to build the connectivity of separated components. The first-order high and low pass filters employ programmable memristors in series with capacitors, which could be easily reconfigured to produce different frequency responses. After passing through the buffers, the 4 frequency band components are modulated by the weighted VMM network according to the equation $V_{Out}(t) = R_x \sum_{j=1}^{n} VBPF_j(t) \cdot W_j$, where $V_{Out}(t)$ and R_x are output voltage and feedback resistance of the trans-impedance amplifier, $VBPF_j(t)$ represents the voltage signal from j-th band pass filter, and W_j is the computing weight (conductance of memristor). If the weight memristor has high conductance, the corresponding frequency band signal is retained. However, low conductance weight memristor could significantly suppress the associated signal.

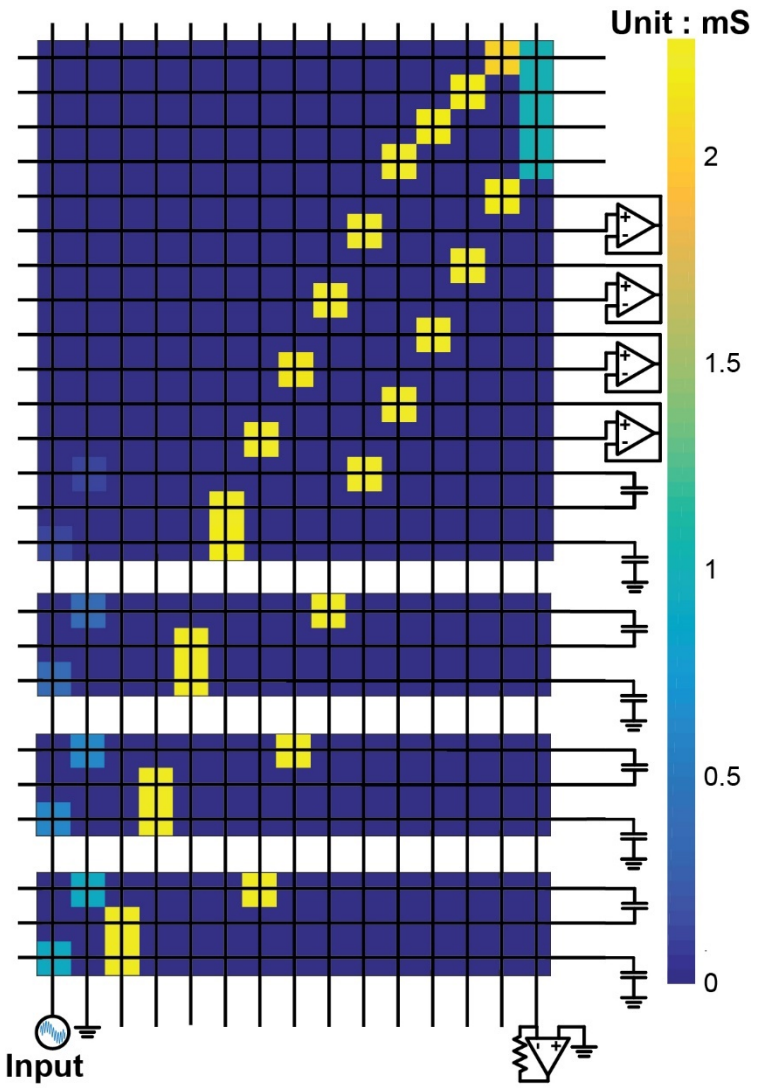


Figure 4.5 : Conductance map of memFPAA audio equalizer. The 24×15 memFPAA subarray configured for the audio equalizer. External TIAs, buffers, and capacitors, as well as input and outputs also shown in the figure. The red dot and the color-map represent the targeting and programmed conductance of the memristor array.

To show the programmable the weight of memristors, we programmed 8 different weight vectors by tuning the weight memristors of the computing core. As shown in Fig. 4.6, 8 different frequency responses have been obtained. A weight vector with all individual weights large (e.g. Fig. 4.6.1) or small (e.g. Fig. 4.6.6) could retain or depress the signals in all bands. While weight vectors with non-uniform weight elements could selectively emphasize signals of interested frequency range as shown in Fig. 4.6.2, 4.6.5.

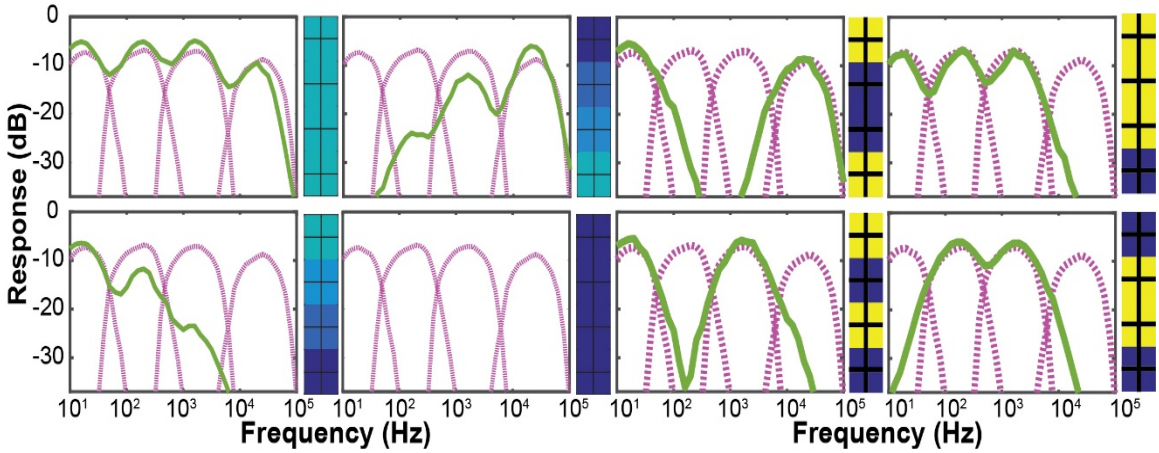


Figure 4.6 : Frequency Response of memFPAA audio equalizer. The frequency responses of the equalizer with different weight vectors. Pink-dash lines are the input frequency response. High conductance weight memristor retains the signal while low conductance memristor suppresses the corresponding signal.

4.5 Frequency-Pattern Recognition

The memristor VMM CAB could serve as a single-layer perceptron neural network to classify vectors in a hyperplane. Here we demonstrate a simple memFPAA based mixed-frequency classifier. As shown in Fig. 4.7, the band pass filters play the role of the basilar membrane of the human hearing system which is selectively resonant to sinusoidal inputs of a particular frequency at different places along the membrane. The signal is then fed to the buffers and output to the single layer perceptron network, with memristor synapses and peak detector neurons. Each post-synaptic neuron, consisting of a trans-impedance amplifier and a series peak detector, receives its own weighted sum of the signals of band pass filters. Since each input pattern is a combination of signals of different frequencies, the neuron associated with synapses with larger weights on the matched frequencies will yield a larger output voltage. This process closely mimics that hair cells sense the vibration of the basilar membrane and transmit signals to cerebral cortex. The inputs comprise 6 different temporal patterns.

Each pattern is a combination of sinusoidal waves of two different frequencies, as illustrated in Fig. 4.8a. A 4×6 memristive synaptic array was pre-trained with measured weights depicted in Fig. 4.8b after programming. The measured peak voltages of the 6 post-synaptic neurons produce unique responses to each input, correctly classifying the input signals of different combination of frequencies. For instance, with the input pattern I4 (i.e. sum of two 1V amplitude sinusoidal waves of frequencies 200Hz and 2kHz), the neurons N1 and N5 observe 200Hz dominant waveforms at the output with peaks ~1V, as they interface with large-weight synapses to the 200Hz filter and but small weight synapses to the 2kHz filter. (See Fig. 4d and e) Similarly, the neurons N2 and N6 output waveforms with 2kHz dominant signal at 1V peak.

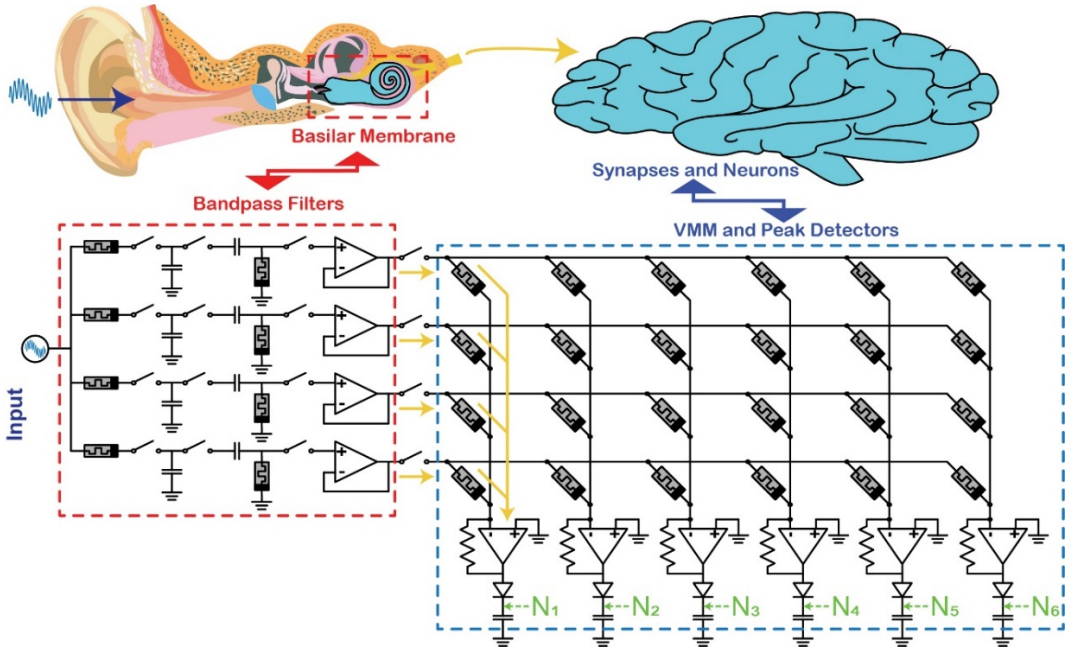


Figure 4.7 : MemFPAA mixed-frequency classifier. Schematic of the classifier circuit consisting of a bank of cascaded high and low pass filters, serving the role of the basilar membrane of the human hearing system, followed by a 4×6 VMM as synaptic array with 6 neurons implemented by TIAs and peak detectors.

The neuron N3 is associated with low weight synapses to both 200Hz and 2kHz filters, resulting almost zero output. On the contrary, neuron N4 observes a waveform

similar to that of the input since synapses to both 200Hz and 2kHz filters are of large weights, producing highest peak ~2V. This proof of principle demonstration of the memFPAA acoustic classifier could be expanded to handle tasks with higher complexity, such as speech recognition.

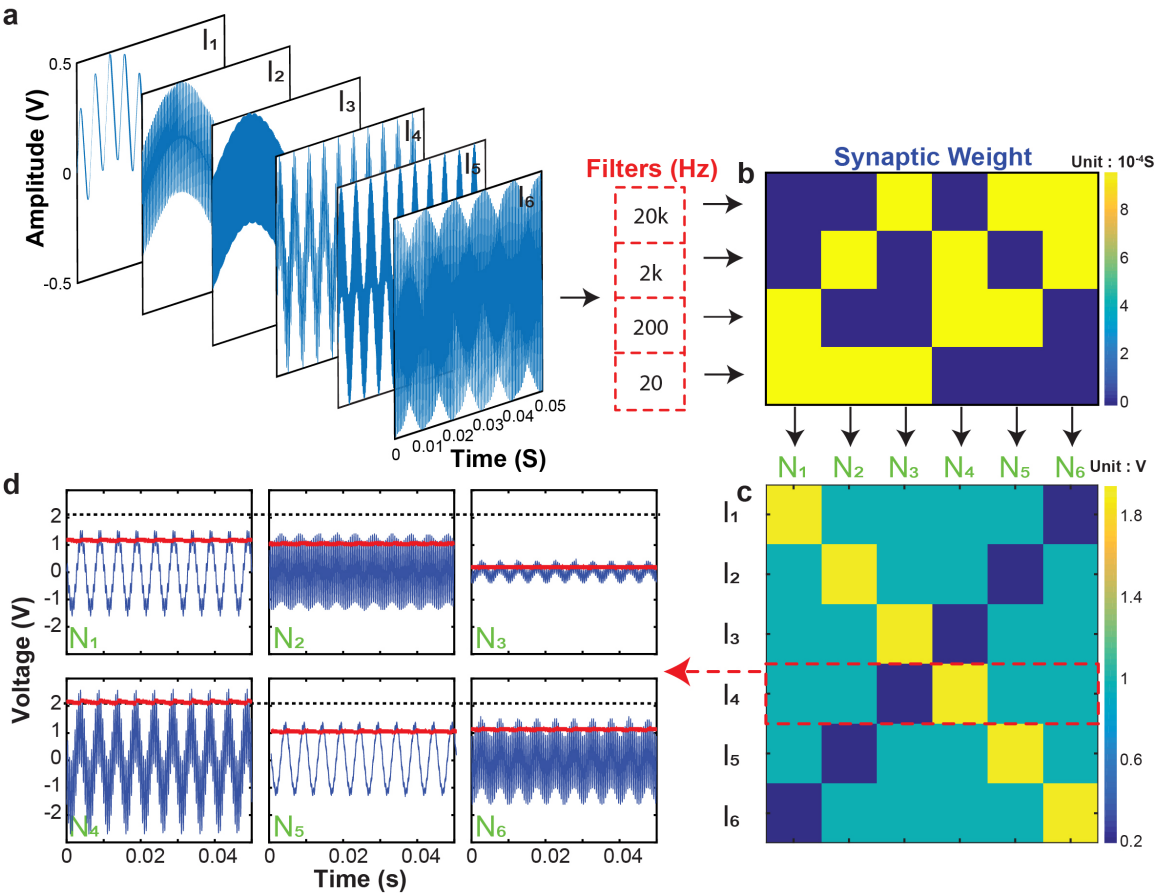


Figure 4.8 : Experimental demonstration of MemFPAA mixed-frequency classifier. a. The inputs comprising 6 different temporal patterns which are combinations of sinusoidal waves of two different frequencies. (From the left, 20Hz and 200Hz, 20Hz and 2kHz, 20Hz and 20kHz, 200Hz and 2kHz, 200Hz and 20kHz, 2kHz and 20kHz.) b. Measured conductance weight of the 4×6 VMM synaptic array. c-d. Measured peak voltages of the post-synaptic neurons to the input patterns. Each individual input pattern is associated with a unique response of artificial neurons. The temporal responses of the 6 output neurons to input pattern I4 is illustrate in d.

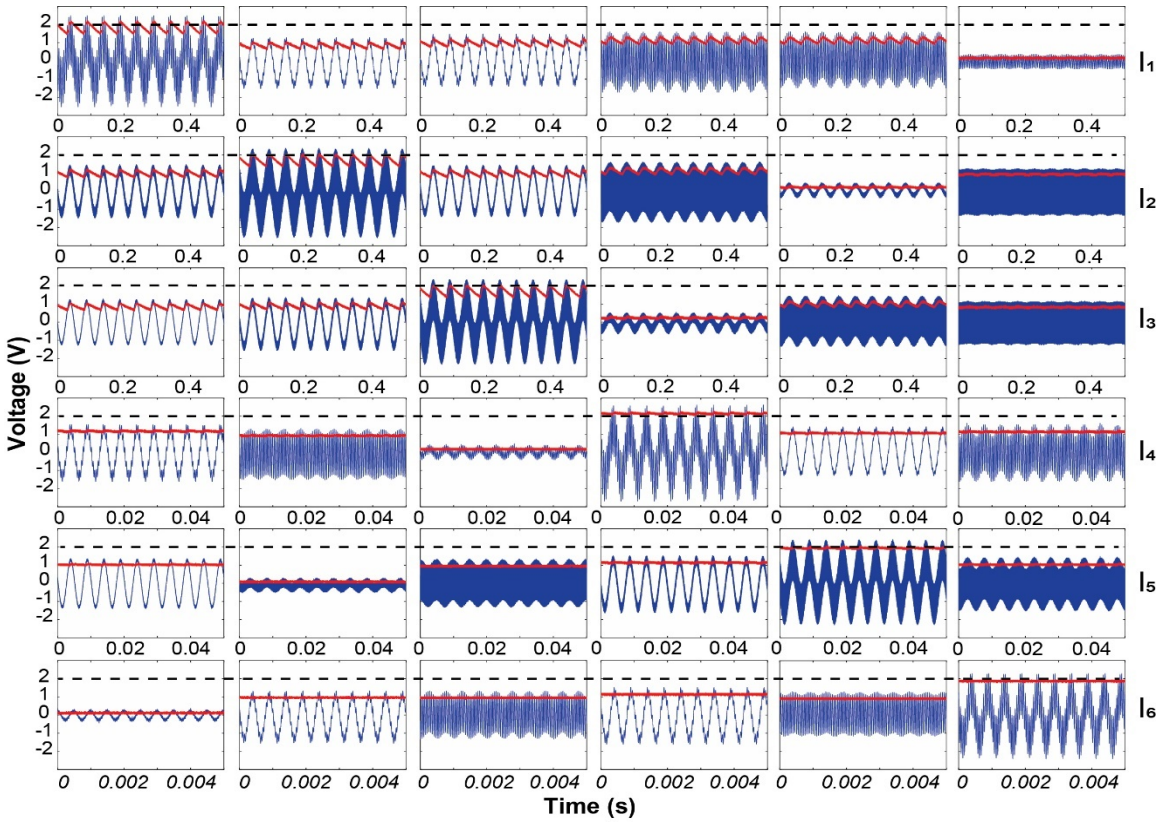


Figure 4.9 : Measured Neuron responses of the MemFPAA mixed-frequency classifier in Fig. 4.8c. The upper schematic shows the synaptic weight matrix. The temporal responses of the 6 output neurons (as indicated by the black arrows) to the 6 input patterns (I1, I2, … , I6) are illustrated in the lower 6×6 panels.

4.6 Summary

A novel implementation of FPAA based on memristor crossbar array has been developed for the first time. Memristor plays various critical roles in the new reconfigurable analog computing structure, not only as switches in the fabric network, but also as reconfigurable resistors for VMMs and other analog units. A variety of computing functions have been experimentally implemented with the memFPAA. Similar to FPGAs, large scale FPAAs can be used either as a general purpose reconfigurable analog circuit or as a platform to quickly prototype analog circuit designs, which can reduce the time of a typical analog circuit design from months to minutes. The demonstrated FPAAs based

on memristor crossbar arrays are expected to significantly advance the development and applications of analog circuits to meet the increasing needs of analog computing.

4.7 Reference

1 Prezioso, M. *et al.* Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* 521, 61-64, (2015).

2 Sheridan, P. M. *et al.* Sparse coding with memristor networks. *Nat Nanotechnol* 12, 784-789, (2017).

3 Gulak, P. G. in *Microelectronics and VLSI, 1995. TENCON '95., IEEE Region 10 International Conference on.* 123-126.

4 Marsh, D. Programmable analogue ICs challenge spice–and–breadboard designs. *EDN Europe*, 30-36, (2001).

5 D'Mello, D. R. & Gulak, P. G. in *Field-Programmable Analog Arrays* 7-34 (Springer, 1998).

6 Fisher, W. A., Fujimoto, R. J. & Smithson, R. C. A programmable analog neural network processor. *Neural Networks, IEEE Transactions on* 2, 222-229, (1991).

7 Austin, K. (Google Patents, 1993).

8 Lee, E. & Gulak, P. Field programmable analogue array based on MOSFET transconductors. *Electron. Lett* 28, 28-29, (1992).

9 Brink, S., Hasler, J. & Wunderlich, R. Adaptive Floating-Gate Circuit Enabled Large-Scale FPAA. (2014).

10 Basu, A. *et al.* A floating-gate-based field-programmable analog array. *Solid-State Circuits, IEEE Journal of* 45, 1781-1794, (2010).

11 Hall, T. S., Hasler, P. & Anderson, D. V. *Field-Programmable Analog Arrays: A Floating—Gate Approach*. (Springer, 2002).

12 Harrison, R. R., Bragg, J. A., Hasler, P., Minch, B. A. & Deweerth, S. P. A CMOS programmable analog memory-cell array using floating-gate circuits. *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on* 48, 4-11, (2001).

13 Smith, P. D., Kucic, M. & Hasler, P. in *Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on.* V-489-V-492 vol. 485 (IEEE).

14 Hall, T. S., Twigg, C. M., Hasler, P. & Anderson, D. V. in *Parallel and Distributed Processing Symposium, 2004. Proceedings. 18th International.* 142 (IEEE).

15 Laiho, M. *et al.* FPAA/Memristor Hybrid Computing Infrastructure. *Circuits and Systems I: Regular Papers, IEEE Transactions on* 62, 906-915, (2015).

16 Laiho, M. *et al.* in *Circuits and Systems (ISCAS), 2014 IEEE International Symposium on.* 2265-2268 (IEEE).

17 Yang, J. J., Strukov, D. B. & Stewart, D. R. Memristive devices for computing. *Nature Nanotech.* 8, 13-24, (2013).

18 Yang, J. J. *et al.* Memristive switching mechanism for metal/oxide/metal nanodevices. *Nature Nano.* 3, 429-433, (2008).

19 Yang, J., Qureshi, M. S., Ribeiro, G. M. & Williams, R. S. (Google Patents, 2014).

20 Jiang, H. *et al.* Sub-10 nm Ta Channel Responsible for Superior Performance of a HfO2 Memristor. 6, 28525, (2016).

21 Hu, M. *et al.* in *Design Automation Conference (DAC), 2016 53nd ACM/EDAC/IEEE.* 1-6 (IEEE).

22 Fujishima, H., Takemoto, Y., Onoye, T. & Shirakawa, I. An architecture of a matrix-vector multiplier dedicated to video decoding and three-dimensional computer graphics. *IEEE transactions on circuits and systems for video technology* 9, 306-314, (1999).

23 Li, C. *et al.* Analogue signal and image processing with large memristor crossbars. *Nature Electronics* **(**In Press**)**, (2018).

24 Bayat, F. M., Alibart, F., Gao, L. & Strukov, D. B. A reconfigurable FIR filter with memristor-based weights. *arXiv preprint arXiv:1608.05445*, (2016).

25 Ali, S., Hassan, A., Hassan, G., Bae, J. & Lee, C. H. in *Optical Fiber Sensors Conference (OFS), 2017 25th.* 1-4 (IEEE).

26 Sozen, H. & Cam, U. in *Electrical and Electronics Engineering (ELECO), 2013 8th International Conference on.* 45-48 (IEEE).

27 Välimäki, V. & Reiss, J. D. All about audio equalization: Solutions and frontiers. *Applied Sciences* **6**, 129, (2016).

28 Kumar, N., Himmelbauer, W., Cauwenberghs, G. & Andreou, A. G. An analog VLSI chip with asynchronous interface for auditory feature extraction. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing* 45, 600-606, (1998).

29 Kesarkar, M. P. in *M. Tech. Credit Seminar Report, Electronic Systems Group, EE. Dept, IIT Bombay.* 1-12.

30 Kim, B. & Pardo, B. in *Machine Learning and Applications (ICMLA), 2014 13th International Conference on.* 495-499 (IEEE).

31 George, S. *et al.* A programmable and configurable mixed-mode FPAA SoC. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 24, 2253-2261, (2016).

CHAPTER 5

CONCLUSION

In this work, about using memristor crossbar arrays for analog computing, the related memristive material system, measurement architecture, and computing applications have been discussed. Memristive devices, as a simple two-terminal passive device, shows scalability, multiple-conductance levels, high write and read speed, low power consumption, and CMOS compatibility, which indicates it will be a promising candidate as the building block of next-generation computing architecture. It also suggests that memristors can, not only be used for data storage, but also for analog computing.

To overcome the disadvantages of using commercial instruments, two measurement systems were designed and realized on the analog computing task. One is based on the synergy of a switching-matrix and B1500A, which provides an easy way to do statistics of a mass of devices from crossbar array. Another system was a dedicated design for parallel computing using 1T1R crossbar array. This can provide multichannel voltage driving and current sensing capabilities simultaneously. It enables demonstrate of analog vector-matrix multiplication based analog applications.

To demonstrate analog computing by using memristors, the vector-matrix multiplication (VMM) was realized on the developed measurement system. Based on the analog VMM, image compression and convolution were successfully implemented on memristor crossbar arrays. During experiments, high performance of devices was achieved, such as 99.8% device yield over 8,000 devices, and at least 64 levels device programming resolution. Although many circuit issues in the analog domain are not

trivial, the system shows reasonable precision, high energy efficiency and computing performance.

Last but not the least, a novel field programmable analog array (FPAA) structure based on memristor crossbar array has been developed for the first time. Memristor plays critical roles in the new reconfigurable analog computing structure, not only as switches in the fabric network, but also as reconfigurable resistors for VMMs and other analog units. A variety of computing functions have been experimentally implemented with the memFPAA, such as first-order pass filter, audio equalizer, and mixed-frequency pattern classifier. The demonstrated FPAAs based on memristor crossbar arrays are expected to significantly advance the development and applications of analog circuits to meet the increasing needs of analog computing.

In summary, analog computing applications based on memristors have been explored in this study. Due to the lack of commercial electrical testing instruments for those emerging devices and crossbar arrays, we have designed and built testing circuits to implement analog and parallel computing operations. With the setup developed in this study, we have successfully demonstrated image processing functions utilizing large memristor crossbar arrays. We further designed and experimentally fabricated the first memristor based FPAA, which was successfully configured for audio equalizer demonstration as an exemplary application of such memristive FPAA (memFPAA).

CHAPTER 6

PUBLICATIONS

1 **Yunning Li,** Z. Wang, P. Yan, P. Lin, C. Li, W. Song, H. Jiang, M. Rao, M. Barnell, Q. Wu, H. Wu, Q. Xia and J. J. Yang. "Memristive Field Programmable Analog Arrays for Analog Computing", *In preparation.*

2 C. Li, D. Belkin, **Yunning Li**, P. Yan, M. Hu, N. Ge, H. Jiang, E. Montgomery, P. Lin, Z. Wang, J. P. Strachan, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang and Q. Xia. "Efficient and self-appetitive in-situ learning in multilayer memristive neural networks", *In preparation.*

3 P. Lin, C. Li, **Yunning Li**, H. Jiang, Z. Wang, W. Song, M. Rao, M. Barnell, Q. Wu, J. J. Yang & Q. Xia, "Parallel Convolutional Kernel Processing Through Three-Dimensional Memristor Circuits", *In preparation.*

4 Z. Wang, S. Joshi, S. Saveliev, W. Song, **Yunning Li**, R. Midya, M. Rao, P. Yan, S. Asapu, Y. Zhuo, H. Jiang, P. Lin, C. Li, J. H. Yoon, N. K. Upadhyay, M. Hu, J. P. Strachan, M. Barnell, Q. Wu, H. Wu, R. S. Williams, Q. Xia and J. J. Yang, "Diffusive memristor artificial neurons for fully memristive neural network" *In preparation.*

5 Z. Wang, M. Rao, J. Zhang, J. Han, P. Lin, **Yunning Li**, C. Li, W. Song, S. Asapu, R. Midya, Y. Zhuo, H. Jiang, J. H. Yoon, N. K. Upadhyay, S. Joshi, M. Hu, J. P. Strachan, M. Barnell, Q. Wu, H. Wu, R. S. Williams, Q. Xia & J. J. Yang, "Capacitive Implementation of Neuronal Dynamics", *In preparation.*

1. 6 M. Hu, C. E. Graves, C. Li, **Yunning Li**, N. Ge, E. Montgomery, N. Davila, H. Jiang, R. S. Williams, J. J. Yang, Q. Xia, and J. P. Strachan, "Memristor-based analog computation and neural network classification with a dot product engine", *Advanced Materials 29, In press (2017)*.

7 C. Li, M. Hu, **Yunning Li**, H. Jiang, N. Ge, E. Montgomery, N. Davila, C. E. Graves, Z. Li, J.P. Stranchan, P. Lin, W. Song, Z. Wang, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang and Q. Xia. "Analog Signal and Image Processing with Large Memristor Crossbars", *Nature Electronics 2017 online.*

8 H. Jiang, D. Belkin, S. E. Savelev, S. Lin, Z. Wang, **Yunning Li**, S. Joshi, R. Midya, C. Li, M. Barnell, Q. Wu, J. J. Yang and Q. Xia. "A Novel True Random Number Generator based on Stochastic Diffusive Memristors", *Nature Communications 2017.*

BIBLIOGRAPHY

1 L. Chua, *et al.* Memristor-The missing circuit element. *IEEE Transactions on Circuit Theory* 18, 507-519 (1971).

2 P. M. Sheridan, *et al.* Sparse coding with memristor networks. *Nat Nanotechnol* 12, 784-789, (2017).

3 D. B. Strukov, *et al.* The missing memristor found. *Nature* 453, 80-83 (2008).

4 J. J. Yang, *et al.* Memristive switching mechanism for metal/oxide/metal nano-devices. *Nat Nanotechnol* 3, 429 (2008).

5 Q. Xia, *et al.* Memristor−CMOS Hybrid Integrated Circuits for Reconfigurable Logic. *Nano Letters* 9, 3640-3645 (2009).

6 J. J. Yang, *et al.* Memristive devices for computing. *Nat Nanotechnol* 8, 13-24 (2013).

7 H. Jiang, *et al.* Sub-10 nm Ta Channel Responsible for Superior Performance of a HfO2 Memristor. *Sci Rep* 6, 28525 (2016).

8 L. Gao, *et al.* Demonstration of Convolution Kernel Operation on Resistive Cross-Point Array. *IEEE Electron Device Letters* 37, 870-873 (2016).

9 M. Hu, *et al.* in *2016 IEEE International Conference on Rebooting Computing (ICRC)*. 1-5 (San Diego, CA, USA, 2016).

10 H. Wu, *et al.* Resistive Random Access Memory for Future Information Processing System. in *Proceedings of the IEEE* PP-99 (2017).

11 B. Chakrabarti, *et al.* A multiply-add engine with monolithically integrated 3D memristor crossbar/CMOS hybrid circuit. *Sci Rep* 7, 42429 (2017).

12 C. Li, *et al.* Analog signal and image processing with large memristor crossbars. *Nature electronics,* (2017 accepted).

13 P. G. Gulak, *et al.* in *Microelectronics and VLSI, 1995. TENCON '95., IEEE Region 10 International Conference on.* 123-126.

14 M. Hu, *et al.* The Dot Product Engine: a memristor-based experimental platform for computational acceleration. (2017, in preparation).

15 H. Kutuk, et al. A field-programmable analog array (FPAA) using switched-capacitor techniques in circuits and Systems ISCAS (1996).

16 M. Prezioso, *et al.* Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* 521, 61-64, (2015).

17 D. Marsh, *et al.* Programmable analogue ICs challenge spice–and–breadboard designs. *EDN Europe*, 30-36, (2001).

18 L. Gao, *et al.* Fully parallel write/read in resistive synaptic array for accelerationg on-chip learning. *Nanotechnology* 26, 455204 (2015).

19 M. Hu, *et al.* in *Proceedings of the 53rd annual Design Automation Conference.* 1-6 (2016).

20 M. Hu, *et at.* The Dot Product Engine: a memristor-based experimental platform for computational acceleration. (2017 in preparation).

21 M. Hu, *et al.* in *Design Automation Conference (DAC), 2016 53nd ACM/EDAC/IEEE.* 1-6 (IEEE).

22 B. Li, *et al.* Exploring the precision for RRAM-based Analog Approximate Computing. *IEEE Design & Test* 33, 51-58 (2016).

23 L. Xia, *et al.* Technological Exploration of RRAM Crossbar Array for Matrix-Vector Multiplication. *Journal of Computer Science and Technology* 31, 3-19 (2016).

24 G. Peng, *et al.* in *The 20th Asia and South Pacific Design Automation Conference.* 106-111, (2015).

25 S. Yu, *et al.* in *Int El devices Meet.* 416-419 (2016).

26 P. M. Sheridan, *et al.* Sparse coding with memristor networks. *Nat Nanotechnol* 12, 784-789, (2017).

27 Y. LeCun, *et al.* Deep Learning. *Nature* 521, 436-444 (2015).

28 M. Hu, *et al.* in *2012 49th ACM/EDAC/IEEE Design Automation Conference (DAC).* 498-503 (2012).

29 S. Park, *et al.* Electronic system with memristive synapses for pattern recognition. *Sci Rep* 5, 10123 (2015).

30 S. Choi, *et al.* Experimental Demonstration of Feature Extraction and Dimensionality Reduction Using Memristor Networks. *Nano Lett 17*, 3113-3118 (2017).

31 C. Li, *et al.* Three-dimensional crossbar arrays of self-rectifying Si/SiO2/Si memristors. *Nat Commun* 8, 15666 (2017).

32 B. Chakrabarti, *et al.* A multiply-add engine with monolithically integrated 3D memristor crossbar/CMOS hybrid circuit. *Sci Rep* 7, 42429 (2017).

33 Prezioso, M. *et al.* Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* 521, 61-64, (2015).

33 Gulak, P. G. in *Microelectronics and VLSI, 1995. TENCON '95., IEEE Region 10 International Conference on.* 123-126.

34 Marsh, D. Programmable analogue ICs challenge spice–and–breadboard designs. *EDN Europe*, 30-36, (2001).

35 D'Mello, D. R. & Gulak, P. G. in *Field-Programmable Analog Arrays* 7-34 (Springer, 1998).

36 Fisher, W. A., Fujimoto, R. J. & Smithson, R. C. A programmable analog neural network processor. *Neural Networks, IEEE Transactions on* 2, 222-229, (1991).

37 Lee, E. & Gulak, P. Field programmable analogue array based on MOSFET transconductors. *Electron. Lett* 28, 28-29, (1992).

38 Brink, S., Hasler, J. & Wunderlich, R. Adaptive Floating-Gate Circuit Enabled Large-Scale FPAA. (2014).

39 Basu, A. *et al.* A floating-gate-based field-programmable analog array. *Solid-State Circuits, IEEE Journal of* 45, 1781-1794, (2010).

40 Hall, T. S., Hasler, P. & Anderson, D. V. *Field-Programmable Analog Arrays: A Floating—Gate Approach*. (Springer, 2002).

41 Harrison, R. R., Bragg, J. A., Hasler, P., Minch, B. A. & Deweerth, S. P. A CMOS programmable analog memory-cell array using floating-gate circuits. *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on* 48, 4-11, (2001).

42 Smith, P. D., Kucic, M. & Hasler, P. in *Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on.* V-489-V-492 vol. 485 (IEEE).

43 Hall, T. S., Twigg, C. M., Hasler, P. & Anderson, D. V. in *Parallel and Distributed Processing Symposium, 2004. Proceedings. 18th International.* 142 (IEEE).

44 Laiho, M. *et al.* FPAA/Memristor Hybrid Computing Infrastructure. *Circuits and Systems I: Regular Papers, IEEE Transactions on* 62, 906-915, (2015).

45 Laiho, M. *et al.* in *Circuits and Systems (ISCAS), 2014 IEEE International Symposium on.* 2265-2268 (IEEE).

46 Yang, J., Qureshi, M. S., Ribeiro, G. M. & Williams, R. S. (Google Patents, 2014).

47 Fujishima, H., Takemoto, Y., Onoye, T. & Shirakawa, I. An architecture of a matrix-vector multiplier dedicated to video decoding and three-dimensional computer

graphics. *IEEE transactions on circuits and systems for video technology* 9, 306-314, (1999).

48 Bayat, F. M., Alibart, F., Gao, L. & Strukov, D. B. A reconfigurable FIR filter with memristor-based weights. *arXiv preprint arXiv:1608.05445*, (2016).

49 Ali, S., Hassan, A., Hassan, G., Bae, J. & Lee, C. H. in *Optical Fiber Sensors Conference (OFS), 2017 25th.* 1-4 (IEEE).

50 Sozen, H. & Cam, U. in *Electrical and Electronics Engineering (ELECO), 2013 8th International Conference on.* 45-48 (IEEE).

51 Välimäki, V. & Reiss, J. D. All about audio equalization: Solutions and frontiers. *Applied Sciences* **6**, 129, (2016).

52 Kumar, N., Himmelbauer, W., Cauwenberghs, G. & Andreou, A. G. An analog VLSI chip with asynchronous interface for auditory feature extraction. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing* 45, 600-606, (1998).

53 Kesarkar, M. P. in *M. Tech. Credit Seminar Report, Electronic Systems Group, EE. Dept, IIT Bombay.* 1-12.

54 Kim, B. & Pardo, B. in *Machine Learning and Applications (ICMLA), 2014 13th International Conference on.* 495-499 (IEEE).

55 George, S. *et al.* A programmable and configurable mixed-mode FPAA SoC. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 24, 2253-2261, (2016).